Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 42 (2017) 1–22

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

# Heterogeneity in generalized reinforcement learning and its relation to cognitive ability

Action editor: Sebastien Helie

Shu-Heng Chen [a,*], Ye-Rong Du [b,a]

[a] AI-ECON Research Center, Department of Economics, National Chengchi University, Taipei, Taiwan
[b] Regional Development Research Center, Taiwan Institute of Economic Research, Taipei, Taiwan

Received 3 April 2016; received in revised form 5 November 2016; accepted 8 November 2016
Available online 19 November 2016

**Abstract**

In this paper, we study the connections between working memory capacity (WMC) and learning in the context of economic guessing games. We apply a generalized version of reinforcement learning, popularly known as the experience-weighted attraction (EWA) learning model, which has a connection to specific cognitive constructs, such as memory decay, the depreciation of past experience, counterfactual thinking, and choice intensity. Through the estimates of the model, we examine behavioral differences among individuals due to different levels of WMC. In accordance with 'Miller's magic number', which is the constraint of working memory capacity, we consider two different sizes (granularities) of strategy space: one is larger (finer) and one is smaller (coarser). We find that constraining the EWA models by using levels (granules) within the limits of working memory allows for a better characterization of the data based on individual differences in WMC. Using this level-reinforcement version of EWA learning, also referred to as the EWA rule learning model, we find that working memory capacity can significantly affect learning behavior. Our likelihood ratio test rejects the null that subjects with high WMC and subjects with low WMC follow the same EWA learning model. In addition, the parameter corresponding to 'counterfactual thinking ability' is found to be reduced when working memory capacity is low.


*Keywords:* Generalized reinforcement learning; Experience-weighted attraction learning; Cognitive ability; Granularity

## 1. Introduction: motivation and literature review

The purpose of this paper is twofold. First, it is a follow-up study to the research on *individual differences in learning* observed in the laboratory of games and markets and characterized by various empirical (parametric) learning models (see Section 1.1). In this literature, learning heterogeneity can be represented by the diversity of the estimates of the models when they are applied to observations associated with different individual subjects or different groups of subjects. Among many possible parametric learning models, *generalized reinforcement learning*, or more popularly known as *experience-weighted attraction (EWA) learning*, is the one strongly motivated by psychology (Camerer & Ho, 1999); hence, it provides us with a natural wonder regarding the possible psychological underpinnings of the observed individual differences in learning. The strength of EWA modeling is that its parameters infer multiple cognitive constructs, such as memory decay, counterfactual thinking, and choice intensity, any of which may be sensitive to individual differences in strategic learning.

In pursuing this line of reasoning, this paper examines two hypotheses related to the effects of working memory capacity on learning, one more general and the other more

* Corresponding author.
*E-mail addresses:* chen.shuheng@gmail.com (S.-H. Chen), yerong.du@gmail.com (Y.-R. Du).

http://dx.doi.org/10.1016/j.cogsys.2016.11.001
1389-0417/

focused. The general one is termed the *working memory hypothesis for individual differences in learning*, and the focused one is termed the *working memory hypothesis for individual differences in counterfactual thinking ability*. The first hypothesis, also referred to as the maintained hypothesis, states that subjects with different WMC do not share the same generalized reinforcement learning model. By pinning down one possible source of the above difference, the second hypothesis further states that subjects with different WMC differ in their counterfactual thinking ability, a specific behavioral parameter of the EWA learning model; in particular, as motivated by the literature to be reviewed in Section 2.2, the hypothesis assumes a positive relationship between WMC and counterfactual thinking ability.

Second, an unintended realization from our work is that the learning model is sensitive to the size (cardinality, granularity) of the set of alternatives (choices, strategies, actions, chunks, and so on). We find that the psychological underpinning can be sensibly identified only when the size (cardinality) is small or, at least, not overwhelmingly large. This constraint may be related to Miller's (1956) concept of limited short-term or working memory capacity (Section 1.2).

In this regard, this paper suggests that the generalized reinforcement learning model can be constrained by reducing its strategy space to the number of items defined by the limits of working memory. Constraining the EWA models by using levels (granules) within the limits of working memory allows for a better characterization of the data based on individual differences in WMC, and by using this constrained version of EWA learning, we find that working memory capacity can significantly affect learning behavior. Our likelihood ratio test rejects the null that subjects with high WMC and subjects with low WMC follow the same EWA learning model; hence, the working memory hypothesis for individual differences in learning is well supported. In addition, under the same constrained version of EWA learning, we find that 'counterfactual thinking ability' is significantly reduced when WMC is moderately low or very low; nevertheless, in the reverse direction, 'counterfactual thinking ability' is not significantly increased with moderately high or very high WMC. Hence, our second hypothesis is only weakly supported.

### 1.1. Individual differences in learning

In recent years, behavioral heterogeneity has not only been identified in game experiments, but has also been related to subjects' cognitive ability. In particular, recent studies have placed emphasis on the correspondence between cognitive ability and strategic sophistication, such as inductive reasoning, iterated dominance, and level-$k$ thinking (Brañas-Garza, García-Muñoz, & González, 2012; Burnham, Cesarini, Johannesson, Lichtenstein, & Wallace, 2009; Devetag & Warglien, 2003; Rydval, Ortmann, & Ostatnicky, 2009; Schnusenberg & Gallo, 2011). Within the extensive literature on those behavioral heterogeneities and their possible cognitive correlates, relatively little research has focused on *learning*, and there have been few attempts to establish a direct relationship between cognitive ability and learning.

This deficit may be partially attributed to the *convergence hypothesis*, i.e., the behavioral heterogeneity observed in initial periods of an experiment, if any, may be temporal after subjects become more experienced. Some early studies involving independent measures of cognitive ability have also shown that even though cognitive ability is correlated with the behavioral heterogeneity in the one-shot guessing game, also known as the beauty contest game (BCG), if the game is played repeatedly this correlation is no longer significant (Burnham et al., 2009; Schnusenberg & Gallo, 2011).

Nevertheless, the convergence property does not guarantee a unique path toward the equilibrium, and one large body of the literature in economics examines the so-called *out-of-equilibrium dynamics*. Hence, the relevance of cognitive ability to individual differences in learning can still be an issue from the perspective of the transition dynamics of the games or markets. By applying individual learning models, several studies have identified individual differences in learning in games (Ho, Wang, & Camerer, 2008) and in markets (Chen & Hsieh, 2011; Hommes, 2011). In addition, there are also experimental studies showing that learning is not independent of cognitive ability (Casari, Ham, & Kagel, 2007).

In the context of a guessing game (beauty contest experiment), Gill and Prowse (2012) found that cognitive ability may positively affect learning in that subjects with higher cognitive ability may learn more actively than subjects with lower cognitive ability and hence, in the end, their performance gap will become even more significant than that at the initial time.

Chen, Du, and Yang (2014) conducted six series of 15- to 20-person beauty contest experiments, and examined the guessing behavior of a set of 108 subjects involved in these experiments. They found a significant correlation between guessing performance and WMC. They also performed regression analysis and found that WMC positively affects reasoning depth. Through a game of up to 10 rounds, the performance gap between the high WMC group and the low WMC group was found to shrink but still existed significantly. They further applied the level-k reasoning model (see Section 3.3) to examine subjects' guessing behavior from round to round. It was found that subjects with high WMC tended to guess with a higher level of reasoning than subjects with low WMC, specifically in the initial periods. Through the analysis of the estimated Markov transition matrix among different levels of reasoning, they further found that the subjects with high WMC had a dynamic behavioral pattern that was different from those with low WMC, which may indicate the possible effect of WMC on learning.

However, neither the level-k reasoning model nor the Markov transition model applied in Chen et al. (2014)

has explicit psychological underpinnings; hence further connections between working memory capacity and learning are not feasible. The mapping between cognitive ability and individual differences in learning requires a formal representation of subjects' learning behavior with a theoretically and empirically sound model; it also requires a repertoire of subjects' personal traits, in this case cognitive ability. Existing studies are either deficient in the former or in the latter. In this paper, we meet the two demands by first applying a well-received learning model, i.e., the generalized reinforcement learning (the EWA learning) model (Section 2.1), to characterize subjects' learning behavior, and then comparing their learning characteristics with their cognitive characteristics which are elicited from a working memory test (Section 3.2). In this way, we can then make sense of the inferred individual differences in learning in light of the subjects' cognitive ability.

More precisely, we shall apply the generalized reinforcement learning (the EWA learning) model to the same data that was employed in Chen et al. (2014), i.e., subjects' behavioral data from the beauty contest experiment (Section 3.1). As we shall see in Section 2, the parameters of the EWA learning model can be linked to specific cognitive constructs, such as memory decay, depreciation of past experience, counterfactual thinking, and choice intensity, which allow for a fine grained analysis of the factors underlying behavioral differences among different WMC groups. We shall see whether the parameter estimates can be related to subjects' working memory in a sensible way, in particular, the parameter normally interpreted as counterfactual thinking (Section 2.2).

We consider two different sizes (granularities) of the set of alternatives: one is large and finer, and one is smaller and coarser. As we shall see later, the results which we have sensitively depend on these settings, and there is a possibility that reinforcement learning may not behave properly when the strategy space is overwhelmingly large. Therefore, before we proceed further, it is also necessary to motivate these settings with related literature developments.

### 1.2. Granularity in reinforcement learning

The second motivation of this paper is to question the *size* of the choice problem to which reinforcement learning is applied. The size here is measured by the number of distinctive choices, options, strategies, actions, chunks, and so on. The reinforcement learning model was originally initiated by psychologists (Bush & Mosteller, 1955), and was later on introduced to economics by Cross (1973), Arthur (1993), and Roth and Erev (1995). When it was used by psychologists, it was applied to deal with stochastic choice when the space of available options, strategies, or actions is rather limited. The typical example is the application of reinforcement learning to the *multi-armed bandit problems*. While one can have many arms (choices) in this problem, it is the *two-armed version* that is most popular. When Arthur (1993) broke the long silence since Cross (1973) first introduced reinforcement learning in economics, the laboratory data which he used were in fact from the *two*-armed bandit experiment conducted by Laval Robillard in 1952–53 at Harvard (Bush & Mosteller, 1955). Even though two arms may be too restrictive, psychologists rarely consider 100 arms. To the best of our knowledge, the largest number of arms that has ever been used in experimental economics is nine (Brenner & Vriend, 2006).

Psychologists seem to be more sensitive to this number of possible choices than economists. A classic work is Miller's famous number *seven* (Miller, 1956). Miller (1956) is a celebrated contribution to psychology in the discussion of *short-term memory capacity* or *working memory capacity*. In this regard, it is about the number of items that an individual can discriminate among or can remember over very short periods of time, say, seconds. Based on a few experiments that he reviewed, Miller concluded that most people can correctly recall about $7 \pm 2$ items. This is the origin of the magic number *seven*.

This paper was overwhelmingly received and motivated many follow-up studies, which even caused an 'evolution' of this magic number, for example, from seven even to four (Cowan, 2001; Mathy & Feldman, 2012). However, its connection to the number of choices in economics became more evident only after the *choice overload hypothesis* was formulated. The hypothesis basically states that "an increase in the number of options to choose from may lead to adverse consequences such as a decrease in the motivation to choose or the satisfaction with the finally chosen option" (Scheibehenne, Greifeneder, & Todd, 2010, p. 73). This hypothesis was first tested by Iyengar and Lepper (2000), and there is a large body of literature emanating from it (Scheibehenne et al., 2010). While there is a constant influx of research trying to distinguish the cases "more is less" from the cases "more is better", the lessons that we have been given from these studies have already become a part of the psychological foundation of public policy when it deals with the design of choice architecture (Iyengar, Huberman, & Jiang, 2004; Thaler & Sunstein, 2008).

The above literature well motivates the fundamental question: to make choice problems susceptible to the analysis of reinforcement learning, need we impose a constraint on the size of the set of alternatives? In fact, the exceedingly large number of alternatives has already been noticed in human-subject experimental games where reinforcement learning is often applied (Chen & Khoroshilov, 2003; Sarin & Vahid, 2004). However, the way in which experimental economists cope with this large size of alternatives is to introduce a *similarity function*, also known as the *neighborhood function*, well used in the machine learning literature. The neighborhood function or the similarity function is to correlate the payoffs of similar strategies up to their degree of closeness or similarity. Nevertheless, just this device alone cannot solve the size problem. To manage

the size problem, we need to have either a direct manipulation of the number of strategies or an indirect reorganization scheme to limit the effective number of choices.

In this paper, we shall propose a direct approach, i.e., to use a coarse granulation to substantially reduce the number of strategies into a niche of Miller's magic numbers. We will then compare the results by applying the EWA learning models to both the original (finer) strategy space and the modified (coarser) strategy space.

## 2. EWA learning

### 2.1. An introduction to the EWA learning model

Learning behavior in experimental games, including BCG, has been widely studied by applying various learning models (Crawford, 1995; Duffy & Nagel, 1997; Ho, Camerer, & Weigelt, 1998; Nagel, 1995; Roth & Erev, 1995; Stahl, 1996, 1998). Among them, a generalized reinforcement learning model, formally known as *experience-weighted attraction (EWA) learning*, was initially proposed by Camerer and Ho (1998, 1999) to encompass two important families of learning models, namely, reinforcement learning and belief learning. These two families may originally be seemingly unrelated, but are now related to each other as special cases in the EWA family. The EWA model has been applied to many experimental games, and the BCG studied in this paper is one of them.

For making this paper self-contained, the EWA learning model is briefly introduced in this section. The EWA learning model as a generalized reinforcement learning model is a kind of stochastic choice model. The stochastic choice model is normally applied to the situation where subjects are repeatedly offered a fixed set of choices (strategies), but the reward for each choice is not fixed, neither certain. This uncertain environment can cause the subject's choice behavior to also be random. In a technical formulation, the random behavior can be represented in a probabilistic fashion, which assigns each choice (strategy) a probability based on its prospect, propensity, or attraction. This choice probability will be updated over time with the subject's experience, and hence learning is encapsulated through this updating mechanism. In the EWA learning model, the choice probability of each strategy is represented by a *logit* function of the strategy's *attraction*, which is determined by the initial attraction of the strategy and is updated through time according to the payoff from choosing that strategy.

To proceed further, let us denote the attraction of strategy $j$ for subject $i$ at time (round) $t$ by $A_i^j(t)$. The appearance of $t$ as part of the notation shows that the strategy attraction is updated over time, a notion suitable for the repeated game. Accordingly, $A_i^j(0)$ is the prior value of the initial attraction before the game starts. The EWA learning model captures the essence of learning through the dynamics of $A_i^j(t)$ (the attraction update). The attraction update mainly depends on two determinants, first, the strategy's own past attraction $A_i^j(t-1)$, and, second, the newly-gained experience of strategy $j$ or its immediate payoff $\pi_i^j(t)$. Let $s_i(t)$ denote player $i$'s choice at time $t$ and $s_{-i}(t) = (s_1(t), \ldots, s_{i-1}(t), s_{i+1}, \ldots, s_n(t))$ denote a strategy combination of all other subjects' strategies at time $t$. Then

$$\pi_i^j(t) = I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t)).$$

where $I(x, y)$ is an indicator function that equals 1 if $x = y$ and 0, otherwise. The indicator function simply shows the possibility that subject $i$ may have no newly-gained experience of strategy $j$ had it not been chosen to activate in time $t$. The update is then the sum of these two determinants.

$$A_i^j(t) = A_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t)). \tag{1}$$

Eq. (1) is a simple version of reinforcement learning. This simple version has, however, been extended by taking into account other psychological factors, such as memory and imagination (Camerer & Ho, 1999; Roth & Erev, 1995). Eq. (2) gives such an extended version with two additional parameters, $\phi$ and $\delta$.

$$A_i^j(t) = \phi \cdot A_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t)). \tag{2}$$

In Eq. (2), the parameter $\phi$ is the normal discount factor which captures the *memory decay*. The parameter $\delta$ dictates whether the payoff of a non-activated strategy (unchosen strategy) can still be simulated or imagined as a result of *counterfactual thinking*, for example, "what might have happened had I not chosen that but this" or "the grass is always greener on the other side". The conventional reinforcement learning model shaped by psychologists only cares about the actual payoff rather than the simulated one; in this case, $\delta = 0$. If, however, the subject does have the power to simulate the what-if scenarios, then part or the whole of the forgone payoff may contribute to attraction updating; in this case, $0 < \delta \leqslant 1$.

The key component of this updating rule (2) is the weighted payoff term

$$[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t)),$$

which captures two basic principles of learning. First, the attractions of chosen strategies $s_i(t)$ are updated by the actual payoff, which means that successful strategies are given more reinforcement and are more likely to be repeated in the subsequent encounters. Behavioral psychologists call this the *law of effect* (Herrnstein, 1970; Thorndike, 1911). Second, the attractions of unchosen strategies are updated by a foregone and hypothetical payoff with the weight $\delta$ ($0 \leqslant \delta \leq 1$). Camerer and Ho (1999) introduce this effect and call it the *law of simulated effect* and rename the former one the *law of actual effect*. In this setting, both chosen and unchosen strategies are 'reinforced' by the payoff that the strategy either yielded or would have yielded.

In addition to the law of simulated effect, Camerer and Ho (1999) also introduce a new parameter called the

*experience weight*, denoted by $N(t)$, to capture how subjects' past experience accumulates over time. While past experience could constantly grow linearly in time, Camerer and Ho (1999) consider a growth function which can generalize reinforcement learning to encompass belief learning models, i.e., to dictate $N(t)$ by a first-order difference equation. It begins with an initial value $N(0)$ and is updated according to

$$N(t) = \rho N(t-1) + 1, \quad t \geqslant 1. \tag{3}$$

The parameter $\rho$ in Eq. (3) controls the speed of the growth of past attractions relative to the current experience, or, in Camerer and Ho (1999)'s expression, "the number of 'ob servation-equivalents' of past experience" relative to one period of current experience. If $\rho$ is small, $N(t)$ will grow at a slower rate or, alternatively put, the past experience will depreciate at a faster rate. On the other hand, if $\rho$ is large, $N(t)$ will grow at a faster rate, which indicates a slower depreciation of past experience. Hence, other things being equal, the higher the $\rho$, the lower the depreciation rate. In this way, the parameter $\rho$ is a discount factor that captures decay in the strength of prior beliefs. The experience weight is then applied to determine the relative importance of the past attraction $A_i^j(t-1)$ and the newly-gained experience $\pi_i^j(t)$ in attraction updating as shown in Eq. (4).

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{N(t)} \tag{4}$$

In general, the attraction $A_i(t)$ is the running total of past attractions, which are constituted by a depreciated experience-weighted past attraction $A_i^j(t-1)$ plus the payoff yielded from period $t$. The probability of choosing strategy $j$ is then determined by its attraction relative to that of other strategies. One mathematical form frequently used to represent this probability function is the logit function, and by the logit function the probability of choosing strategy $j$ at time $t+1$, $P_i^j(t+1)$, is

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{l=1}^{m} e^{\lambda \cdot A_i^l(t)}}, \tag{5}$$

where $m$ denotes the number of choices, and the parameter $\lambda$ ($0 \leqslant \lambda < \infty$) is known as the *intensity of choice* in the literature (Brock & Hommes, 1997). By Eq. (5), the choice is random but is biased toward the strategies with higher attraction values except in the following two extreme cases: $\lambda = 0$ and $\lambda = \infty$. For the former case ($\lambda = 0$), the choice becomes uniformly random, and the attraction value of strategies does not play a role. For the latter case ($\lambda = \infty$), the choice becomes deterministic, and only the strategy with the highest attraction value will be selected.

With Eqs. (4) and (5) it can be shown that the familiar reinforcement learning model and the belief-based learning model are both special cases of the EWA learning model if some parameters of Eq. (4) are properly restricted. As we have already seen above, the family of choice reinforcement models corresponds to the case where $\delta = 0$, and the family of belief learning models corresponds to the case where $\delta = 1$. Between the two extremes, there exist many other learning models characterized by different values of $\delta$, varying between 0 and 1. These models indicate that the law of simulated effect, which distinguishes between the two extremes, can be just *a matter of degree*. Therefore, by encompassing the two extremes, EWA learning allows each subject to have his/her own degree (capability) of engaging in "simulation" or "imagination" or "counterfactual thinking".

### 2.2. Cognitive ability and counterfactual thinking

Through the way in which the EWA learning model is presented, we can see its connections to a number of cognitive constructs, memory decay ($\phi$), depreciation of past experience ($\rho, N(0)$), counterfactual thinking ($\delta$), and choice intensity ($\lambda$). Among these four facets of cognitive constructs, not all of them have a clear relation to working memory capacity. For example, to the best of our knowledge, there is no theory which can inform us of the relationship between WMC and choice intensity ($\delta$). Of course, we can empirically examine their relationship in our BCG experiment, as we shall do in Section 4, but there is no theory known to us to suggest an expectation and to form a hypothesis. The other three parameters, $\phi, \rho$ and $N(0)$, to some extent are all memory-related. One might also expect to find a relation between WMC and these parameters. However, as well explained in Camerer and Ho (1999, p. 839) there is no unique interpretation for the behavior captured by these parameters; for example, they could be interpreted as either *naturally forgetting* or *deliberately forgetting (discounting old experience) when the environment is changing*. Depending on the interpretation adopted, the relation between cognitive capacity and these memory-related parameters can be either positive or negative. This ambivalent relation does not suggest that there is a clear hypothesis to maintain. Among the five, the only one which has a clear psychological underpinning is counterfactual thinking ($\delta$).

Psychologists have hypothesized the relationship between working memory capacity and counterfactual thinking ability for more than a decade (Byrne, 2005). To think counterfactually, one has to keep both the real world and the fictive alternative in mind and therefore it expects the role of working memory capacity (Byrne, 2005, 2016). For instance, low-memory-span individuals are more prone to bias during counterfactual judgment when they are under memory loads (Goldinger, Kleider, Azuma, & Beike, 2003). In addition to the mock-jury decision, studies in psycholinguistics have suggested that the counterfactual context may be more cognitively demanding and requires increased working memory capacity to process (Ferguson, 2012; Kulakova & Nieuwland, 2016; Urrutia, de Vega, & Bastiaansen, 2012). Camille et al. (2004) found that participants with orbitofrontal lesions, while perform-

ing a gambling task, reported no regret and also failed to anticipate the possible negative consequences of their choices. This indicates that individuals with prefrontal cortex damage were less likely to produce counterfactual reasoning.

The recent progress in empirical studies based on heterogeneous learning models normally shows that the estimated parameters differ among subjects (Broseta, 2000; Camerer & Ho, 1998, 1999; Chen & Hsieh, 2011; Cheung & Friedman, 1997; Gill & Prowse, 2012; Ho et al., 2008; Stahl, 2000). However, few of them actually went further to ask what we may learn from these 'inferred individual differences in learning'. This 'silence' is particularly intriguing when the estimated models originate from psychology and are shaped by psychological concepts. Given their psychological underpinning, it makes sense to probe the possible connections between these 'inferred individual differences in learning' and personal traits, such as working memory capacity. Reinforcement learning or its generalized version, EWA learning, is a perfect example. Its parameter $\delta$ has been motivated as an ability to engage in counterfactual thinking, imagination, or simulation (Camerer & Ho, 1999), all of which require cognitive resources. Due to individual differences this task is easier and less costly for some subjects than others. Subjects who are more resourceful can perform this task more easily, and hence may tend to do so at a lower cost, whereas subjects who are less resourceful cannot afford to do so. Therefore, their difference in cognitive ability may be revealed through their behavior (observations) in the lab and may be further captured by the applied statistics. If so, we shall be able to examine the psychological underpinning of the parameter $\delta$ by empirically testing its connection with subjects' WMC.

## 3. Experiments, data and estimation

In this section, we shall first give a description of the experiment (Section 3.1) and the working memory test (Section 3.2) that were conducted to generate the data. We then describe two kinds of empirical EWA learning models used in this paper (Section 3.4). They are distinguished by the size (granularity) of the strategy space: one larger (finer) and one smaller (coarser). The latter is closely tied to the level-$k$ reasoning and classification (Section 3.3). Finally, these models are estimated using maximum likelihood estimation (Section 3.5).

### 3.1. Beauty contest experiment

The experiment consisted of 6 sessions and there were 15–20 subjects in each session, there being a total of 108 subjects involved. The subjects were required to complete both a repeated beauty contest game and then a working memory test. All experiments were conducted in the Experimental Economics Laboratory (EEL) at National Chengchi University from October 2009 to August 2010. Experiments were announced on the NCCU EEL web site[1] and on the part-time job board in PPT, one of the most popular bulletin board systems in Taiwan. Subjects were required to register through the NCCU EEL Registration System[2] and sign up for our experiments. After signing up for one experiment, subjects immediately received an e-mail for confirmation.

The BCG was conducted by means of a z-tree (Fischbacher, 2007). For each period, subjects were required to select an *integer* number between [0, 100], and competed with all of the others in the session. The subjects were informed that the prize was given to the one whose guess number was closest to the *target number*, denoted by $\tau$, which was calculated by averaging all guesses and then multiplying the result by a factor $p = 2/3$. After collecting all subjects' guesses, the screen would display feedback information regarding the target number, the subject's chosen number, profit and accumulated profit. The BCG was repeated 10 times and it took about 60 min to finish. In addition to a fixed show-up fee of NT$125, we also provided a prize for the winner in each period of NT$100, and the prize was to be evenly split if there was more than one winner. The instruction manual can be found in Chen et al. (2014).

### 3.2. Working memory task

The task we used for eliciting working memory capacity (WMC) was developed by Lewandowsky, Oberauer, Yang, and Ecker (2010). This task includes 5 tests, a backward digit-span task (Dspan), a spatial short-term memory test (SSTM), a memory updating task (MU), a sentence-span task (SentSpan), and an operation-span task (OpsSpan). The WMC task was always conducted after finishing the beauty contest experiment and was administered in the order of DSpan, SSTM, MU, SentSpan, and OpsSpan for all subjects. It took about 90 min to finish all 5 tests. NT$200 was paid to those subjects who completed all five sets. The score for each test was calculated and then normalized by the mean and standard deviation of the scores derived from our Experimental Subject Database (ESD), which included 740 subjects completing the same task.

Then a single measure of WMC was derived by averaging these five normalized scores. Although each of these five tests alone is a way of measuring WMC, a battery consisting of heterogeneous indicators is required to reduce the test-specific variance. Lewandowsky et al. (2010) found that, through a structural equation model analysis, the tests including SSTM, MU, SentSpan and OpsSpan have substantial loadings on a single latent WMC factor. These results lend support to our derivation of the WMC score. Compared to the other 4 tests, the Dspan test is more simple without the demand for processing or relational inte-

[1] see http://eel.nccu.edu.tw/.
[2] see http://eel.nccu.edu.tw/Registration/.

gration. However, since the storage capacity of memory measured by Dspan also contributes to the common WMC factor (although weaker), we also include it in the battery of WMC. The descriptions of the stimuli, design, and procedure of all five tests are given in Appendix A.

### 3.3. Characterization of reasoning levels

Subjects' behavior in the beauty contest experiment is frequently associated with level-$k$ reasoning (Nagel, 1995). A simple way to identify subjects' reasoning levels-$k$, known as the Cournot myopic best response algorithm, was proposed by Nagel (1995). Denote player $i$'s guess in period $t$ by $g_i(t)$ and the number of players in session $j$ by $n_j$. For $t = 1$ and $p = \frac{2}{3}$, player $i$ is classified *exactly* as

$$\begin{cases} k_0 \text{ (level 0)}, & \text{if } g_i(1) = 50, \\ k_1 \text{ (level 1)}, & \text{if } g_i(1) = 50 \times p \approx 33.33, \\ k_2 \text{ (level 2)}, & \text{if } g_i(1) = 50 \times p^2 \approx 22.22, \\ k_3 \text{ (level 3)}, & \text{if } g_i(1) = 50 \times p^3 \approx 14.81. \end{cases} \tag{6}$$

Although subjects may not choose these critical values, making a guess closer to any one of them could be *roughly* considered as belonging to the same level. Therefore, we divide $[0, 100]$ into several adjacent intervals corresponding to different levels. For $t > 1$, the subjects are given information about the previous target number. It is plausible that they make the best response to the behavior in the previous period by assuming others to be the same. To sum up, formally, subject $i$ will be classified as level $d$ in period $t$, and denoted by $d_i(t)$, if

$$g_i(t) \in \begin{cases} (m(t-1)p^{d+0.5}, m(t-1)], & \text{if } d = 0, \\ (m(t-1)p^{d+0.5}, m(t-1)p^{d-0.5}], & \text{if } d \neq 0, \end{cases} \tag{7}$$

where

$$m(t-1) = \begin{cases} 50. & \text{if } t = 1, \\ \frac{1}{n_j}\sum_{i=1}^{n_j} g_i(t-1), & \text{if } t > 1. \end{cases}$$

In Eq. (7), the upper limit of the level $d = 0$ is bounded from the right side by $m(t-1)$, instead of $m(t-1)p^{-0.5}$. Regarding this asymmetry, Nagel (1995) indicates that the results, for the first period, would not change if a symmetric bound were to be taken instead (Nagel, 1995, p. 1317). For the later periods, it is pointed out that the chosen numbers tend to be below the mean of the previous period (Nagel, 1995, p. 1320). To make our results comparable with those of Nagel (1995), the same asymmetric bound is taken in our analysis.

Nagel (1995) found that $d = 0$, 1, 2 and 3 can identify approximately 80% or more guesses. The subjects with guesses larger than the upper limit of the level zero ($d = 0$) are grouped into "$d < 0$". Similarly, the subjects with guesses smaller than the lower limit of the level 3 ($d = 3$) are grouped into "$d > 3$". The subjects are finally categorized into 6 classes: $d < 0, d = 0$, 1, 2, 3 and $d > 3$. Normally the target number $\tau$ and hence $m(t)$ will decrease with time; therefore, the Intervals constantly updated according to Eq. (7) are expected to shrink over time. Since all the guessing numbers are integers, it becomes possible that some levels $d$ may contain no single integer.

### 3.4. Empirical EWA learning model

In this subsection, we shall describe how to prepare the EWA learning model so that it becomes an empirical model. Altogether we shall introduce five different empirical EWA learning models, denoted by Models I, II, III, IV, and V, respectively. Model I is introduced by Camerer and Ho (1999), and Models II and III are introduced by Camerer, Ho, and Chong (2002). All these three models are based on the set of alternatives being the integer numbers from 0 to 100. In other words, the cardinality of the strategy space is 101. The subsequent two models, Models IV and V, are developed based on the set of alternatives being the reasoning level (Section 3.3), and the size of the strategy space is 6. Models IV and V are basically the small-size equivalent of Models I and II, respectively. We will then apply these five different EWA learning models to structurally characterize the learning behaviors in repeated BCGs.

#### 3.4.1. Number reinforcement: Models I, II, and III

*Model I.* Camerer and Ho (1999) provide the first attempt to estimate the EWA parameters with the data of repeated BCGs. To estimate the model, since the general structure described in Section 2 is not sufficient, Camerer and Ho (1999) provide further details of the design necessary to carry out the empirical work, specifically, in the context of the BCG. The strategy space of the game is considered to be the interval [0, 100] with only integers allowed. These 101 strategies are initially endowed with attractions $A^j(0)$ $(j = 0, \ldots, 100)$, which are then reinforced over time. If we treat all $A^j(0)$s as model parameters, then there are overwhelmingly 101 parameters to be calibrated.

To alleviate this 'curse of dimensionality', they first grouped these 101 numbers into 10 non-overlapping intervals with consecutive numbers, such as [0, 9], [10, 19],..., and [90, 100]; they then assumed that the initial attraction $A^j(0)$ was the same within each interval. Therefore, the first design detail is that the 101 initial attractions are reduced to only 10, denoted by $A^1(0), A^2(0), \ldots, A^{10}(0)$. This simplification will be referred to as "Assumption A1" throughout our paper. Next, to facilitate a proper calculation of the reinforcement being attributed to each strategy, they further made two additional assumptions. First, subjects know the winning number $w = \arg\min_{g_i}\{|g_i - \tau|\}$ (Assumption A2); second, they treat the target number as being exogenous to their own guesses (Assumption A3). A remark is needed here. Virtually speaking, Assumption A2 can neither be applied to Camerer and Ho (1999)'s nor to our designs of the experiment. This is because sub-

jects in these experiments were only informed of the target number, but not the winning number. They can know the winning number of a specific round only if they happened to be the winner of that round.

Denote the distance between the winning number and the target number as $e = |\tau - w|$, and also denote the prize for each round as $n\pi$ where $n$ is the number of winners in a particular round. By assumptions A2 and A3, we can easily define the reinforcement intervals without introducing additional parameters. First, all subjects reinforce numbers in the intervals $(\tau - e, \tau + e)$ by $\delta n\pi$ (Fig. 1, panels (a) and (b)). This is because none of them actually chose the numbers within this interval; however, through their counterfactual thinking, they know that, had they done so, they would have won a prize of $n\pi$. As a result, based on our earlier discussion of the EWA learning model (Section 2), the non-chosen strategy will be reinforced by this imaginary payoff, as only a $\delta$-proportion of the true payoff.

Second, the winners reinforce what they chose, which is one of the boundary numbers, either $\tau - e$ or $\tau + e$, by $\pi$, and reinforce the other boundary number, which they did not choose, by an imaginary reinforcement of $\delta\pi$ (Fig. 1, panel (a)). Third, losers reinforce both boundary numbers $\tau - e$ and $\tau + e$, again, by an imaginary reinforcement of $(\delta n\pi)/(n+1)$ (Fig. 1, panel (b)). The denominator in the previous division is $n+1$ and not $n$, because had the subject actually chosen the number, we would have one additional subject to share the prize.

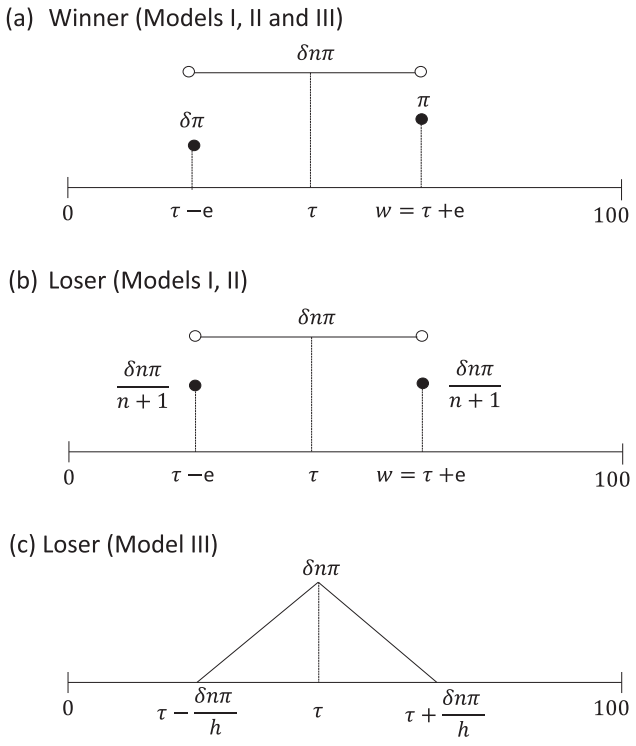



Fig. 1. Calculation of Forgone Payoffs. The above figures summarize how foregone payoffs are determined (see the text for the details); $\delta, \tau$ and $w$ refer to the 'counterfactual thinking ability', the target number and the winning number, respectively. Based on the assumption whether the winning number, $w$, is publicly known (Assumption A2 in the text), two different forgone payoffs are considered for losers, as shown in (b) and (c).

*Model II.* In Camerer et al. (2002) the assumptions A1 and A2 were removed and replaced by others. First, instead of treating initial attractions $A^j(0)$ as parameters and simultaneously estimating them with other EWA parameters, Camerer et al. (2002) empirically obtained $A^j(0)$ from the choice data in the first period. Formally, they recovered initial attractions by the following system of equations

$$\frac{e^{\lambda \cdot A^j(0)}}{\sum_{l=1}^{10} e^{\lambda \cdot A^l(0)}} = f^j, \quad j = 1, \ldots, 10. \tag{8}$$

where $f^j$ is the observed fraction of total activations in the first period that involves strategy $j$. In the system of equations (8), we have 11 unknowns, including ten initial attractions ($A^j(0)$s) plus the parameter $\lambda$, but there are only ten knowns. Hence, the values of the 11 unknowns are underidentified. To solve this underidentification problem, they added one more known by setting the initial attraction of the strategy with the lowest $f^j$ to 0, i.e.,

$$A^{j^*}(0) = 0, \text{ where } j^* = \arg\min_j \{f_j\}. \tag{9}$$

*Model III.* Camerer et al. (2002) also remove the unrealistic assumption A2 and introduce an additional parameter $h$, to be estimated, to describe how losers compute foregone payoffs. We assume that they reinforce numbers in the interval $\left[\tau - \frac{\delta n\pi}{h}, \tau + \frac{\delta n\pi}{h}\right]$. The amount of reinforcement is $\delta n\pi$ at the target number $\tau$, which is the maximum. In departing from $\tau$, the amount of reinforcement decreases at a rate of $h$. The foregone payoff for the losers to be reinforced will be of a triangular form (Fig. 1, panel (c)). Based on a similarity concern, not Assumption A2, Camerer et al. (2002) also assign this parameter $h$ to the winners and assume that they reinforce the numbers in the interval $\left[\tau - e, \tau - e - \frac{\delta n\pi}{h}\right]$ and $\left[\tau + e, \tau + e + \frac{\delta n\pi}{h}\right]$ with a similar triangular form of reinforcement amount, if there is only one winner. However, since Assumption A2 still applies to winners, we therefore decide to stick to our original setting.

Model III differs from the first two models by also taking into account the *similarity effect*. Due to this effect, not only is the target number rewarded by the payoff $n\pi$, but all other neighboring numbers are also rewarded in proportion to their distance from the target number $\frac{n\pi}{h}$. This idea is referred to as *local experimentation* in Roth and Erev (1995). The parameter $h$ is used to control the radius of the neighborhood.

### 3.4.2. Level reinforcement: Models IV and V

In both versions of the EWA model, i.e., Camerer and Ho (1999) and Camerer et al. (2002), subjects are assumed to be able to initialize the attractions of 101 strategies, store all of them in memory, and update them over time. As we have reviewed in Section 1.2, this cognitive task, to some extent, may literally be beyond what a human's limited memory capacity can handle. Therefore, an alternative

approach is to assume that subjects take coarser granules to group numbers and use *interval* instead of *number* as the basic unit of the strategy space.

There are many possible ways to do this; for example, it can be handled by a fuzzy set with flexible linguistic expressions, such as high, medium, and low. The numeric correspondences of these linguistic values are more fluid and can adapt with subjects' dynamically changing perceptions during the game. Hence, 'low' at the beginning of the game may be different from 'low' at the end of the game. In fact, this time-dependent virtue is well captured by the levels associated with level-$k$ reasoning; as described in Eqs. (6) and (7), these levels are time-variant, and can approximately match the perceptions of subjects on the linguistic expressions, high, medium, low, and so on. Therefore, we use the six levels determined by level-$k$ reasoning, namely, $s^j \in \{d < 0, d = 0, d = 1, d = 2, d = 3, d > 3\}$, as the alternative strategy space, and assume that subjects apply and reinforce level-$k$ *rules* instead of 101 numbers. To be distinguished from the previous three models which are based on number reinforcement, we shall also call this alternative, based on level reinforcement, *EWA rule learning*, and the original one *EWA number learning*. In this setting, subjects are only required to initialize, maintain, and update 6 attractions $A_i^j(t)$, where $j \in \{1, 2, 3, 4, 5, 6\}$. We define the payoff function as follows,

$$\pi_i(s_i^j, s_{-i}(t)) = \begin{cases} n\pi, & \text{if } s_i^j = \text{target} - d, \\ 0, & \text{if } s_i^j \neq \text{target} - d, \end{cases}$$

where target-$d$ denotes the level at which the target number is located. Notice that, in our definition, a level is an interval and it may include several numbers.

By Eq. (2), the amount of reinforcement for target-$d$ will be $n\pi$ if it is chosen, and will be $\delta n\pi$ if it is not chosen. Here, we still assume that subjects can infer the level target-$d$ from the known target number. This assumption is similar to Assumption A2, but may be weaker, because in using EWA rule learning we have already implicitly assumed that all subjects are aware that other subjects are learning simultaneously. Their perceived width of each interval is also updated over time, and hence, to some extent, may be approximately close to those given by Eq. (7).

#### 3.4.3. A sum-up

Table 1 summarizes the five empirical EWA learning models applied in this study. As discussed above, these five models differ in their granularities of the strategy space (Table 1, column 2), and also in their set of parameters to be estimated, as shown in the last block of columns under the heading "Parameters". For granularity (the second column), there are two granulations being considered: Models I-III have 101 strategies (numbers), and Models IV and V have only 6 strategies (levels). As to the parameters, all five models share a common set of the five parameters (the third column) to be estimated. Models I and IV also include the initial attractions $A^j(0)$ ($j = 0, 1, 2, \ldots, 101$ for Model I, or $j = 1, 2, \ldots, 6$ for Model IV) as additional parameters, but other models simply use the empirical initial fractions to derive the calibrated ones through Eq. (8). Finally, the radius parameter $h$ (controlling the radius of the target number), as an alternative to Assumption A2, is only imposed in Model III. The other four models do not need this parameter because Models I and II rely on Assumption A2, whereas Models IV and V rest upon a coarser granulation of the strategy space, which already encapsulates a notion of the neighborhood of the target number. To sum up, through Models I-III, we first try to 'replicate' what Camerer and his colleagues have done before (Camerer & Ho, 1999; Camerer et al., 2002) using our data, and take these results as benchmarks to be compared with our proposed models (Models IV-V).

### 3.5. Estimation strategy

To estimate the above models, attractions are first transformed to the choice probability by the logit function, which is given by

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{l=1}^{m} e^{\lambda \cdot A_i^l(t)}}$$

where $m$ denotes the number of choices, and $m = 101$ in Camerer's EWA number learning and $m = 6$ in our EWA rule learning. Define player $i$'s initial attractions as a vector $\mathbf{A}_i(0) \equiv (A_i^1(0), A_i^2(0), \ldots, A_i^m(0))$. We assume a representative agent, so $\mathbf{A}_i(0) = \mathbf{A}(0), \forall i$. The number of subjects is denoted by $I$. Multiplying $I$ by the number of rounds ($T$) for each session, we have a total of $T \cdot I$ observations.

Table 1
Summary of the five EWA learning models.

| Model | Strategy Representation | Parameters | | |
|---|---|---|---|---|
| | | $N(0), \phi, \rho, \delta, \lambda$ | $A^j(0)$ | $h$ |
| I | Number | Estimated | Estimated | NA, Assumption A2 |
| II | Number | Estimated | Initialized | NA, Assumption A2 |
| III | Number | Estimated | Initialized | Estimated |
| IV | Level-$k$ rule | Estimated | Estimated | NA, Coarse Granulation |
| V | Level-$k$ rule | Estimated | Initialized | NA, Coarse Granulation |

Model IV (V) is the equivalent of Model I (II) in a more granular strategy space.

Denote it by $M$. Then the log-likelihood function $LL(\mathbf{A}(\mathbf{0}), N(0), \phi, \rho, \delta, \lambda)$, is

$$LL(\mathbf{A}(\mathbf{0}), N(0), \phi, \rho, \delta, \lambda)$$
$$= \sum_{t=1}^{T}\sum_{i=1}^{I} \ln\left(\sum_{j=1}^{m} I(s_i^j, s_i(t)) \cdot P_i^j(t)\right)$$
$$= \sum_{t=1}^{T}\sum_{i=1}^{I} \ln\left(\sum_{j=1}^{m} I(s_i^j, s_i(t)) \cdot \frac{e^{\lambda \cdot A_i^j(t-1)}}{\sum_{l=1}^{m} e^{\lambda \cdot A_i^l(t-1)}}\right). \quad (10)$$

By following Camerer and Ho (1999), we impose the following restrictions on the parameter space:

$$0 \leqslant A^j(0) \leqslant T \cdot n\pi, \quad \forall j$$
$$\phi > 0,$$
$$0 \leqslant \rho \leqslant 1,$$
$$0 \leqslant \delta \leqslant 1,$$
$$0 \leqslant N(0) \leqslant \frac{1}{1-\rho},$$
$$\lambda > 0.$$

According to Camerer and Ho (1999), these restrictions are based on the following considerations. First, $A^j(0)$ is restricted to be less than or equal to the difference between the maximum and minimum payoffs through the entire game so that attractions are given the same range as payoffs. This restriction allows $N(0)$ to play a role of weighting between initial attractions and payoffs and excludes the possibility of being a scaling factor which puts the attractions and payoffs on the same scale.[3] Second, although both $\rho$ and $\phi$ are discount factors or decay rates and are expected to share a common range between zero and one, $\phi$ is not bounded from the above, which allows us to detect the problem of misspecification especially when the resulting estimates of $\phi$ are above one.[4] Third, the parameter $\delta$ should be equal to or less than one in order to represent a relative weight on the foregone payoff compared to the weight on the actual payoff ($\delta + (1-\delta) = 1$). Fourth, the restriction imposed on $N(0)$ is to ensure that the experience weights increase over time since, based on the first-order difference equation of the experience weight (3), the steady state of $N(t)$ is $1/(1-\rho)$. Finally, $\lambda$ as described in Eq. (5), should positively associate attractions with choice probabilities.

The parameters are first estimated by the choice data of all 108 subjects in all ten rounds. In this way, our results can be compared with the original work done by Camerer and Ho (1999) and a modified version by Camerer et al. (2002). We then separate our data into two groups, high WMC and low WMC (see also Appendix B), and obtain two sets of parameter estimates, to see how cognitive ability affects learning behaviors. To have the results reported here, we actually tried several numerical nonlinear global optimization methods available in Mathematica, including simulated annealing, Nelder-Mead, random search, and differential evolution, to maximize the likelihood function (10). When a specific method derived superior results, we further explored some options of this method, such as the number of search points, number of random seeds, and post process for local search, to avoid reporting local optima.

To be able to conduct model comparison, we followed Camerer and Ho (1999) to calculate and present AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). The former is defined as $LL - k$ and the latter is defined as $LL - (2/k)log(M)$, where $k$ is the number of degrees of freedom and $M$ is the sample size. To conduct legitimate model comparisons using AIC and BIC, it is required that the same dependent variable be used for all candidate models. In this sense, it is not possible to compare the accuracy of number reinforcement models (Models I-III) and level reinforcement (Models IV and V) together since the former models have guessing numbers as the dependent variable and the latter have the choosing level as the dependent variable.

## 4. Results

### 4.1. Number reinforcement models

#### 4.1.1. Models I, II and III

We first estimate the first version of Camerer's EWA number learning model (Camerer & Ho, 1999). The results are given in Table 2. Table 2, column 2, gives the estimates of Model I by pooling all 108 subjects' data together. The last column is the corresponding estimates obtained by Camerer and Ho (1999). By comparing these two columns item by item, we can see some similarities and differences.

First, for the parameters which control the decay rate of memory or experience, such as $\phi, \rho$, and $N(0)$, the results between the two are close. Second, as to the parameters of initial attractions $\mathbf{A}(0)$, direct comparisons of initial attractions are not feasible because they are bounded by $T \cdot n\pi$, which may differ among different experiments. To be specific, the prize for each period, $n\pi$, was either 100 new Taiwan dollars in our case or 3.5 Singapore dollars in Camerer and Ho (1999)'s case. Despite this being so, their relative magnitudes are exactly the same. Both have a peak at $A^1(0)$, then decline constantly until they reach $A^8(0)$, and bounce back again to the end. Third, as to another parameter pertinent to choice making, $\lambda$, our results are rather low as compared to those of Camerer and Ho (1999). Finally, for the parameter of counterfactual thinking ability, $\delta$, we obtain a value of zero, which is also different from their positive estimate, 0.232.

To see whether there is a relation between working memory capacity and $\delta$, we divide our subjects into four subgroups by their WMCs, namely, the top one-third

[3] See Camerer and Ho (1999), p. 846, footNotes 24 and 25.

[4] For the discussion on the range of $\phi$, see Camerer and Ho (1999), pp. 864 and 869.

Table 2
Parameter estimates of EWA number learning: Model I.

| | Model I | | | | | Camerer and Ho (1999) |
|---|---|---|---|---|---|---|
| Parameters | All Subjects | WMC > $P_{67}$ | WMC > mean | WMC < mean | WMC < $P_{33}$ | All Subjects |
| *Initial values* | | | | | | |
| $A^1(0)$ | 1000.000 | 650.995 | 999.653 | 1000.000 | 1000.000 | 3.348 |
| $A^2(0)$ | 843.453 | 603.302 | 853.518 | 850.305 | 827.940 | 3.311 |
| $A^3(0)$ | 609.908 | 556.747 | 721.618 | 499.077 | 485.753 | 3.301 |
| $A^4(0)$ | 602.262 | 533.672 | 635.722 | 607.027 | 595.728 | 3.269 |
| $A^5(0)$ | 409.08 | 467.274 | 492.692 | 378.091 | 327.518 | 3.227 |
| $A^6(0)$ | 385.661 | 491.440 | 464.772 | 357.925 | 312.050 | 3.180 |
| $A^7(0)$ | 293.776 | 412.806 | 374.657 | 279.607 | 292.980 | 3.052 |
| $A^8(0)$ | 0.000 | 335.744 | 0.346 | 0.000 | 0.000 | 2.192 |
| $A^9(0)$ | 137.497 | 319.480 | 333.387 | 25.0297 | 28.356 | 2.871 |
| $A^{10}(0)$ | 392.352 | 463.211 | 547.117 | 272.644 | 123.005 | 3.060 |
| $N(0)$ | 12.890 | 2.773 | 2.485 | 13.578 | 15.165 | 16.815 |
| *Decay parameters* | | | | | | |
| $\phi$ | 1.236 | 1.231 | 1.222 | 1.255 | 1.257 | 1.330 |
| $\rho$ | 0.922 | 0.639 | 0.598 | 0.926 | 0.934 | 0.941 |
| *Imagination factor* | | | | | | |
| $\delta$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.232 |
| *Payoff sensitivity* | | | | | | |
| $\lambda$ | 0.003 | 0.010 | 0.003 | 0.002 | 0.002 | 2.579 |
| *Log-likelihood* | | | | | | |
| LL | −3707.17 | −1227.00 | −1747.83 | −1949.78 | −1507.71 | −5878.20 |
| *Information Criteria* | | | | | | |
| AIC | −3722.17 | −1242.00 | −1762.83 | −1964.78 | −1522.71 | −5893.20 |
| BIC | −3759.56 | −1271.35 | −1794.73 | −1997.24 | −1553.19 | −5932.38 |
| *Sample size* | | | | | | |
| M | 1080 | 370 | 520 | 560 | 430 | 1372 |

($WMC > P_{67}$), above average, below average, and the bottom one-third ($WMC < P_{33}$), and separately estimate the EWA model for each group (see Appendix B). The results are presented in columns 3–6, Table 2. From these columns, some differences and similarities between the high WMC groups ($P_{67}$ and 'above average') and the low WMC groups ('below average' and $P_{33}$) are also observed.

First, they differ in the experience-decaying parameters, $\hat{\rho}$, and $\widehat{N(0)}$. The high WMC groups of subjects have a lower $\hat{\rho}$, indicating that they depreciate their past experience faster than the low WMC groups (see the model description in Section 2). In addition, the high WMC groups have lower initial values of $\widehat{N(0)}$, indicating that they learn faster than low WMC groups, because they attach lower weights to lagged attractions. Second, despite the quantitative difference in their initial attractions ($\mathbf{A}(0)$), the two groups share a very similar pattern, which demonstrates a greater initial interest in the lower values. $A^j(0)$ declines from the beginning ($j = 0$) all the way down and bounces back when $j$ approaches the endpoint ($j = 10$). While a serious behavioral interpretation of this pattern is difficult, this inverted $J$ pattern is basically what we observe for all $\mathbf{A}(\mathbf{0})$, including the one in Camerer and Ho (1999). Third, maybe the greatest commonality shared by these groups is the counterfactual thinking ability. Interestingly enough, $\hat{\delta}$ is consistently zero from the low WMC groups to the high WMC groups. From this result, there is no observed relation between cognitive capacity and $\delta$.

#### 4.1.2. Models II and III

We then move to estimate the second and third versions of Camerer and Ho's EWA learning model (Camerer et al., 2002). Following Camerer et al. (2002) we consider two modifications of the original EWA learning model. In Model II, we initialize $\mathbf{A}(0)$ by the corresponding empirical choice distribution over all strategies in the first period. In addition to that, in Model III, we introduce an additional parameter $h$ to replace the unrealistic assumption that the winning number is known. The parameter estimates of Models II and III are given in Table 3. The last two columns of Table 3 also show the results of Camerer et al. (2002) in which both of the two aforementioned modifications were taken into account, i.e., Model III. Notice that Camerer et al. (2002) recruit both experienced subjects and inexperienced subjects, and their results are separately estimated. While cognitive capacity is not identical to experience, to make a rough comparison, we also present our results by dividing the subjects into the high group (above average) and the low group (below average).

First, let us look at the memory and experience decaying parameters. In Model II, for $\phi, \rho$, and $N(0)$, we find that

Table 3
Parameter estimates of EWA number learning: Models II and III.

| Parameters | Model II[a] | | Model III[b] | | Camerer et al. (2002) | |
|---|---|---|---|---|---|---|
| | WMC > mean | WMC < mean | WMC > mean | WMC < mean | Experienced | Inexperienced |
| $N(0)$ | 0.000 | 0.000 | 0.000 | 0.924 | –[c] | –[c] |
| $\phi$ | 0.701 | 0.683 | 0.881 | 0.685 | 0.22 | 0.000 |
| $\rho$ | 0.000 | 0.000 | 0.725 | 0.388 | 0.000 | 0.000 |
| $\delta$ | 0.436 | 0.598 | 0.354 | 0.899 | 0.99 | 0.90 |
| $\lambda$ | 0.040 | 0.030 | 0.101 | 0.038 | –[c] | –[c] |
| $h$ | – | – | 0.727 | 1.955 | 0.11 | 0.13 |
| LL | −2333.17 | −2526.92 | −1974.96 | −2068.30 | −2128.88 | −2155.09 |
| AIC | −2338.17 | −2531.92 | −1980.96 | −2074.30 | –[c] | –[c] |
| BIC | −2348.80 | −2542.74 | −1993.72 | −2087.28 | –[c] | –[c] |
| M | 520 | 560 | 520 | 560 | 1372 | 1372 |

[a] In Model II, initial attractions $A^j(0)$ are initialized by first period data.
[b] In Model III, initial attractions $A^j(0)$ are initialized by first period data and an additional parameter $h$ is introduced to replace the unrealistic assumption.
[c] Camerer et al. (2002) neither reported the parameter estimates of $N(0)$ and $\lambda$ nor information criteria AIC and BIC in their paper.

subjects in the high and low WMC groups did not perform substantially differently; their numerical difference in $\hat{\phi}$ is negligible. In Model III, these two groups are in stark contrast in all these parameters. The high WMC group exhibits a lower $\widehat{N(0)}$, indicating a rapid initial rate of learning, but a larger $\rho$, indicating a slower depreciation of past experience; the latter is exactly opposite to what we have learned from Model I. The high WMC group also has a slower decaying memory than the low WMC. This inequality is consistent with the one in Camerer et al. (2002) if we match their experienced group with our high WMC group and their inexperienced group with our low WMC group.

Second, regarding the intensity of choice, $\lambda$, the $\hat{\lambda}$ of both models slightly increases as compared to that of Model I. Nevertheless, the inequality direction remains unchanged. In all three models, the high WMC groups tend to have higher $\hat{\lambda}$s than the low WMC groups.[5] The value of $\lambda$ is not reported in Camerer et al. (2002), and hence a further comparison is not available.

Third, maybe the most significant change with the technical modification(s) is $\hat{\delta}$. In Model I, it is consistently zero, but now it is moderately high for both groups and for both models. Furthermore, $\hat{\delta}$ is different between the two groups and that inequality direction is consistent in both models, indicating that $\delta$ is higher for the low WMC group. This result may contradict the way in which we motivate the connection between $\delta$ and cognitive capability (see Section 2.2). In comparison with Camerer et al. (2002), while our $\hat{\delta}$ is positive, it is much lower than theirs, which are both close to one. In addition, in Camerer et al. (2002) experienced subjects were found to have a higher $\hat{\delta}$ than inexperienced subjects. This inequality direction is expected if $\delta$, as the ability to engage in counterfactual thinking, can be related to experience.

Finally, the newly-added parameter $h$ in our data is much greater than the one in Camerer et al. (2002), which suggests that when calculating the foregone payoff, our subjects applied a triangular form with a narrower base (a smaller neighborhood), hence the foregone payoff quickly goes toward zero once the strategy (the guessing number) gets away from the target number, and subjects with a lower WMC have an even smaller neighborhood than subjects with a higher WMC.

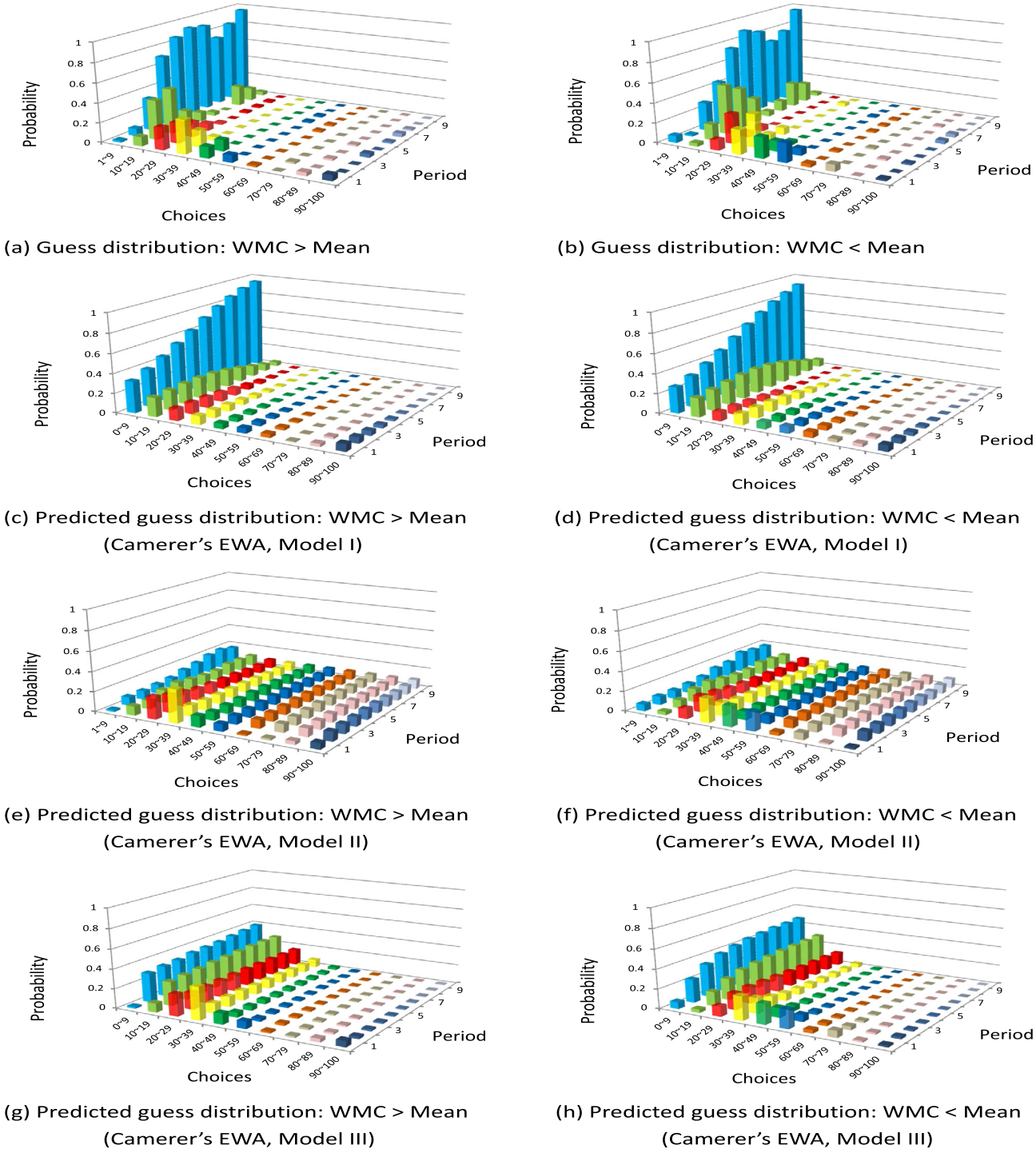
In sum, Camerer et al. (2002) provide us with some fine details about the heterogeneity of the behavioral parameters. By roughly treating cognitive ability in parallel to experience, we find that in both memory and radius parameters we have consistency in the inequality direction; however, in regard to our counterfactual thinking parameter, $\delta$, our observed inequality directions are just the opposite of those in Camerer et al. (2002).

#### 4.1.3. Predicted choice probabilities

It would be interesting to know how well the previous three versions of the EWA learning model perform in terms of actually predicting (fitting) the empirical distribution of the strategy choice. To do so, in Fig. 2, the top panel, we present the evolution of the empirical distribution of the strategy choice from period one to period ten. To make a comparison, in the upper middle, lower middle, and bottom panels, we also present the evolution of the predicted probability of each strategy being chosen (the predicted frequencies of guessing numbers) using Models I, II, and III, respectively. As before, we separate the observations into a high WMC group (higher than average) and a low WMC group (lower than average). They are displayed on the left and the right of each panel, respectively.

Let us first look at the predictions made by Models II and III (the lower middle panel and the bottom panel).

[5] Regarding the relation between $\lambda$ and performance, Chen and Hsieh (2011) is the only experimental study known to us. In their asset market experiments with 120 subjects, $\hat{\lambda}$ ranges from a minimum of 0.01 to a maximum of 42,746. Their empirical results indicate that subjects with a greater $\lambda$ tend to perform better than they would do otherwise.




(a) Guess distribution: WMC > Mean

(b) Guess distribution: WMC < Mean

(c) Predicted guess distribution: WMC > Mean (Camerer's EWA, Model I)

(d) Predicted guess distribution: WMC < Mean (Camerer's EWA, Model I)

(e) Predicted guess distribution: WMC > Mean (Camerer's EWA, Model II)

(f) Predicted guess distribution: WMC < Mean (Camerer's EWA, Model II)

(g) Predicted guess distribution: WMC > Mean (Camerer's EWA, Model III)

(h) Predicted guess distribution: WMC < Mean (Camerer's EWA, Model III)

Fig. 2. Guess distribution and predicted guess distribution.

Although EWA learning seems to predict a tendency toward convergence to zero (the unique Nash equilibrium), its speed of convergence is too slow to mimic experimental data. Next, let us look at the prediction made by Model I (the upper middle panel). Camerer and Ho (1999) comment on its performance as follows:

> None of these models captures the nature of learning well. The reinforcement and one-segment EWA models simply pretend that the first period is like later periods and inflate initial attractions to gradually reproduce the later-period data. (Camerer & Ho, 1999, p. 866)

This comment applies well to our performance using Model I. As we can see, the first-period prediction made by Model I is substantially different from the actual choice distribution. The entire predicted distribution of the first period is like that of the later periods, and, with $\hat{\phi}$ larger

than one (Table 2), it gradually converges to or reproduces the last-period distribution. As pointed out in Camerer and Ho (1999), $\hat{\phi}$ being larger than one is an indication of model misspecification.

In fact, Camerer and Ho (1999) horse raced various learning models including choice reinforcement, belief-based, and EWA learning. They found that, although the EWA learning model does do better in fitting other games such as the constant-sum game and the median-action coordination game, when applied to the beauty contest game, it does not have an equally good degree of fitness. One remedy to this problem, as suggested by Camerer and Ho (1999), is to consider learning when players sophisticatedly realize that other players are learning as well. Sophistication is central in BCGs for producing level-$k$ reasoning and it has been put into practice in Camerer et al. (2002).

Here, we do not follow the sophisticated version of EWA learning; instead, we turn to a different direction, which may have been largely ignored in the literature, i.e., the size of strategy space. Based on our concern as discussed in Section 1.2, there is a possibility that (general) reinforcement learning may not behave properly when the strategy space is overwhelmingly large. In fact, the two experiments in which EWA learning performs well in Camerer and Ho (1999) both have a much smaller strategy space; it is seven for the median action game, and four or six for the constant sum game. These sizes are all in the range of 'magic numbers'. Therefore, in this paper, we would like to apply the same EWA learning to a smaller (coarser) strategy space to see whether simple EWA learning can properly function.

### 4.2. Level reinforcement models

#### 4.2.1. Models IV and V

We redefined the strategy space from 101 guessing numbers to 6 reasoning levels. Based on the discussion in Section 1.2, we consider that few strategies are more plausible than many strategies for subjects to distinguish, process, and hence reinforce them. In this case, levels of reasoning are independently calculated and their attractions are directly reinforced. The two versions of EWA rule learning proposed in Section 3.4.2 were estimated, and the results of the estimates are given in Table 4 (Model IV) and Table 5 (Model V).

By taking a glimpse at these two tables, we can find that the parameter estimates are not sensitive to whether $A^j(0)$ is estimated by MLE (Eq. (10)) or is calibrated by empirical distribution (Eq. (8)). To make a quick comparison with the EWA (number) learning model, we also demonstrate the empirical choice probability sided with the predicted choice probability based on Models IV and V in Fig. 3, as a juxtaposition of Fig. 2.

First of all, if we look at the first-period prediction, we can see that both models have matched that of the empirical distribution quite closely. Even though the initial distribution of Model IV is derived by the maximum likelihood estimator (Eq. (10)) rather than by the direct imposition (Eq. (8)), it still catches well the mode at '$d = 1$' for the high WMC group (panel (c), Fig. 3) and the mode at '$d = 0$' for the low WMC group (panel (d), Fig. 3). Second, compared to Model I ((c) and (d), Fig. 2), nither of these two models generates the distribution of the later periods through inflating or deflating the initial distribution. Instead, from period to period, we can see the shift of the mode from one level to a different level; besides, the ups and downs of major levels (spikes) in the empirical distribution are also well represented by the predictions of Models IV and V. Finally, apart from the level distribution, we also find that $\hat{\phi}$ uniformly lies between zero and one (Tables 4 and 5); hence, the early misspecification found in Model I no longer exists. With the above features, we tend to conclude that the simple EWA learning model can still perform well even in the BCG as long as the size of the strategy space is consistent with the human's mental constraint.

Now, let us come back to Tables 4 and 5. Notice that we add one more high and one more low group to these tables, i.e., the top one fourth ($WMC > P_{75}$) and the bottom one fourth ($WMC < P_{25}$) (see Appendix B). Hence, we have three high WMC groups and three low WMC groups. Models IV and V are applied to each group to derive the parameter estimates of the respective group. In light of the subjects' WMC and better model quality, can we make good sense of the estimates obtained?

There are two ways in which we can form the comparison, one using only the symmetric group and one using any two of the six groups. The former restricts the comparison to within the three symmetric pairs: ('WMC > mean', 'WMC < mean'), ('WMC > $P_{66}$', 'WMC < $P_{33}$'), and ('WMC > $P_{75}$', 'WMC < $P_{25}$'); the farther away from the mean, the further sharper the contrast. The symmetric comparison serves this purpose: it examines whether cognitive ability affects the learning behavior in more and more contrasting frames. Alternatively, any higher WMC group can be compared to any lower WMC group; in this way we can see whether the influence of cognitive ability is monotonic. To make our presentation easier, we shall use the 'weak sense' to refer to the comparisons of symmetric pairs only, and the 'strong sense' to refer to comparisons of any pairs, and unless it is mentioned we mean only in the 'weak sense'. We shall begin by looking at the structure of the numerical differences, and leave the former test of some of these differences to the next section.

Given that the parameter estimates of models IV and V are quite similar, the following results on the effect of working memory capacity generally apply to both models. We begin with the experience and memory decay factors, $\widehat{N(0)}$, $\hat{\rho}$, and $\hat{\phi}$. As we can see from the tables, subjects with high WMC tend to learn faster (a lower $\widehat{N(0)}$), are more sensitive to recent experience (a lower $\hat{\rho}$), and have a slower

Table 4
Model parameter estimates of EWA rule learning: Model IV.

| Parameters | All Subjects | WMC > $P_{75}$ | WMC > $P_{67}$ | WMC > mean | WMC < mean | WMC < $P_{33}$ | WMC < $P_{25}$ |
|---|---|---|---|---|---|---|---|
| $A^1(0)$ | 532.286 | 533.361 | 518.615 | 500.240 | 533.144 | 553.022 | 513.975 |
| $A^2(0)$ | 545.508 | 493.314 | 468.395 | 475.770 | 549.254 | 559.275 | 510.991 |
| $A^3(0)$ | 576.813 | 789.98 | 664.918 | 600.983 | 551.464 | 574.821 | 511.407 |
| $A^4(0)$ | 549.469 | 791.015 | 678.365 | 573.574 | 524.537 | 524.52 | 504.834 |
| $A^5(0)$ | 462.426 | 646.955 | 549.271 | 451.966 | 468.705 | 430.489 | 486.738 |
| $A^6(0)$ | 466.480 | 609.044 | 512.761 | 416.919 | 484.360 | 474.161 | 497.761 |
| $N(0)$ | 0.464 | 0.490 | 0.584 | 0.210 | 1.131 | 1.287 | 4.918 |
| $\phi$ | 0.889 | 0.889 | 0.914 | 0.861 | 0.848 | 0.665 | 0.820 |
| $\rho$ | 0.464 | 0.000 | 0.000 | 0.000 | 0.711 | 0.223 | 0.949 |
| $\delta$ | 0.535 | 0.411 | 0.484 | 0.578 | 0.472 | 0.405 | 0.314 |
| $\lambda$ | 0.016 | 0.010 | 0.009 | 0.011 | 0.027 | 0.017 | 0.090 |
| LL | −1756.01 | −480.35 | −562.99 | −815.47 | −927.29 | −712.28 | −566.63 |
| AIC | −1767.01 | −491.35 | −573.99 | −826.47 | −938.29 | −723.28 | −577.63 |
| BIC | −1794.43 | −512.08 | −595.51 | −849.87 | −962.09 | −745.63 | −598.68 |
| M | 1080 | 320 | 370 | 520 | 560 | 430 | 340 |

In Model IV, initial attractions $A^j(0)$ are estimated by all data.

Table 5
Model parameter estimates of EWA rule learning: Model V.

| Parameters | All Subjects | WMC > $P_{75}$ | WMC > $P_{67}$ | WMC > mean | WMC < mean | WMC < $P_{33}$ | WMC < $P_{25}$ |
|---|---|---|---|---|---|---|---|
| $N(0)$ | 0.328 | 0.333 | 0.178 | 0.083 | 0.706 | 14.068 | 9.171 |
| $\phi$ | 0.891 | 0.906 | 0.939 | 0.864 | 0.849 | 0.677 | 0.493 |
| $\rho$ | 0.465 | 0.260 | 0.362 | 0.114 | 0.675 | 1.000 | 0.964 |
| $\delta$ | 0.545 | 0.464 | 0.558 | 0.588 | 0.489 | 0.400 | 0.277 |
| $\lambda$ | 0.016 | 0.014 | 0.015 | 0.012 | 0.025 | 0.255 | 0.181 |
| LL | −1756.95 | −483.01 | −564.93 | −815.66 | −929.06 | −713.72 | −569.03 |
| AIC | −1761.95 | −488.01 | −569.93 | −820.66 | −934.06 | −718.72 | −574.03 |
| BIC | −1774.41 | −497.43 | −579.72 | −831.29 | −944.88 | −728.88 | -583.60 |
| M | 1080 | 320 | 370 | 520 | 560 | 430 | 340 |

In Model V, initial attractions $A^j(0)$ are initialized by first period data.

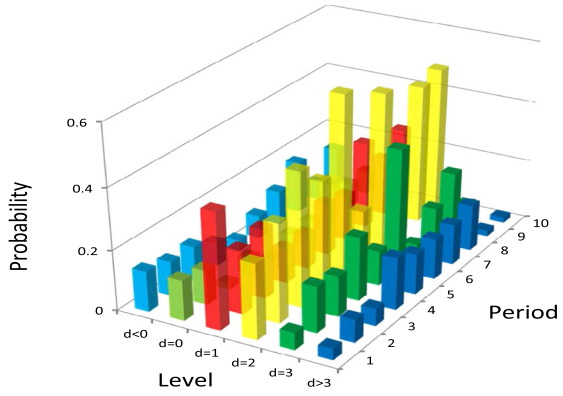
memory decay rate (a larger $\hat{\phi}$). This is basically consistent with what we learn from Model I and Camerer et al. (2002) (Table 3). Second, regarding the intensity of choice ($\lambda$), we have a result that is totally contradictory to the previous models: subjects with high WMC now have a lower $\hat{\lambda}$ than subjects with low WMC. Before seeing this result, we have a strong tendency to suspect that $\lambda$ is positively related to performance and is positively associated with cognitive ability (also see footnote 5). However, the results here have added some uncertainties regarding this relationship. Finally, it is the counterfactual thinking ability ($\delta$). Among all symmetric pairs, subjects with high WMC tend to have a higher $\hat{\delta}$ than subjects with low WMC. This result overthrows the findings of Model I (no relationship) and Models II and III (a negative relationship). However, due to relatively superior model quality, we tend to attach a higher weight to this finding, and tend to treat this result more seriously.

All the qualitative results shown above are in the weak sense, i.e., valid only for the symmetric pairs. If we just read the parameter estimates row by row from one end to the other end, we can quickly see that we cannot find any evidence of the monotonic relation, i.e., the strong sense. Take $\delta$ as an example. The top one-fourth of subjects in WMC have a $\hat{\delta}$ higher than the bottom one-fourth of subjects, but compared to the top one-third or top one-half, their $\hat{\delta}$ is lower. Interestingly enough, it is even lower than the bottom one-half. To ascertain whether this indicates a non-linear effect of WMC on $\delta$, a more rigorous statistical test is needed.

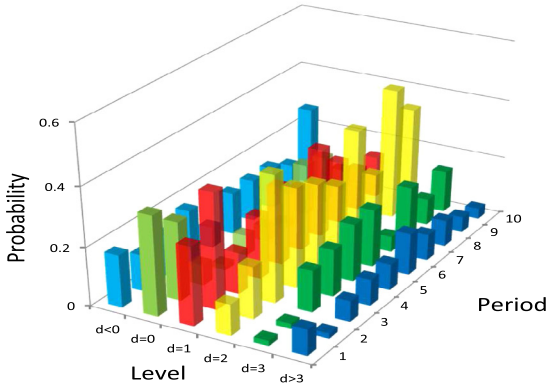
#### 4.2.2. Likelihood ratio tests

In this section, we shall apply the likelihood ratio (LR) test to provide a statistical treatment of the significance of working memory capacity (WMC) to learning. The likelihood ratio test is carried out at two levels which, from general to specific, correspond to two questions. First, at a more general level, we inquire whether the learning behavior, characterized by the parameters of the EWA learning model, differs between the high WMC group and the low WMC group. Second, at a specific level, we ask whether the counterfactual thinking ability, characterized by the parameter $\delta$, differs between the two groups.

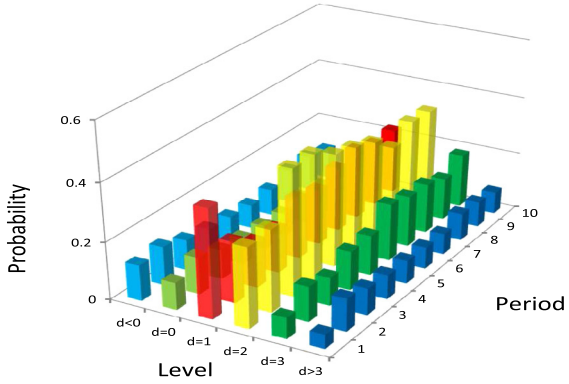
To proceed, let us denote the vector of the estimates of the respective parameters by $\hat{\boldsymbol{\theta}} = (\widehat{N(0)}, \hat{\phi}, \hat{\rho}, \hat{\delta}, \hat{\lambda})$. Furthermore, we shall use the superscripts '$h$' and '$l$' to distinguish the vectors associated with high and low WMC groups, i.e., $\hat{\boldsymbol{\theta}}^h$ and $\hat{\boldsymbol{\theta}}^l$. Let $LL(\cdot)$ denote the log-likelihood function, and



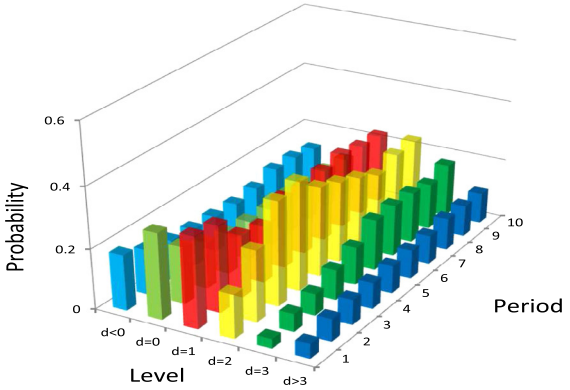

(a) Level distribution: WMC > Mean



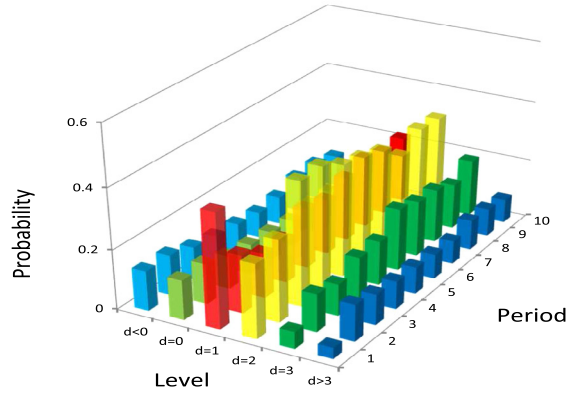

(b) Level distribution: WMC < Mean



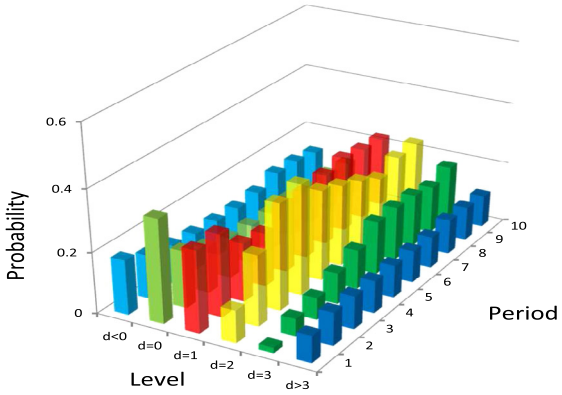

(c) Predicted level distribution: WMC > Mean (EWA rule, Model IV)




(d) Predicted level distribution: WMC < Mean (EWA rule, Model IV)




(e) Predicted level distribution: WMC > Mean (EWA rule, Model V)




(f) Predicted level distribution: WMC < Mean (EWA rule, Model V)

Fig. 3. Level distribution and predicted level distribution.

we also distinguish its association with the two groups using the same superscripts, i.e., $LL^h(\cdot)$ and $LL^l(\cdot)$. When the function $LL(\cdot)$ is evaluated under a given vector of estimates, $\hat{\boldsymbol{\theta}}$, it is written as $LL(\hat{\boldsymbol{\theta}})$. We shall use the high WMC group as the reference. In this case, $\hat{\boldsymbol{\theta}}^h$ can be regarded as the *unrestricted* MLE estimate, and $LL^h(\hat{\boldsymbol{\theta}}^h)$ is the unrestricted likelihood of the high-WMC observations.

To conduct the general test of the difference in learning behavior between high and low WMC groups, we simply replace $\hat{\boldsymbol{\theta}}^h$ with $\hat{\boldsymbol{\theta}}^l$ to $LL^h(\cdot)$ as if restricting parameters $\theta = \hat{\theta}^l$ (the null hypothesis) and obtain the corresponding $LL^h(\hat{\boldsymbol{\theta}}^l)$. This defines the restricted likelihood of the high-WMC observations. The likelihood ratio test statistic, denoted by $LR$, for the null can be written as

$$LR = -2(LL^h(\hat{\boldsymbol{\theta}}^l) - LL^h(\hat{\boldsymbol{\theta}}^h)).$$

It is known that under the null the test statistic $LR$ follows a chi-square distribution with $r$ degrees of freedom, which is the number of restrictions; in our case, $r = 11$ for Model IV, and $r = 5$ for model V.

On the other hand, to conduct the specific test of the difference in counterfactual thinking ability ($\delta$) between the high and low WMC groups, we replaced $\hat{\delta}^h$ with $\hat{\delta}^l$ into $LL^h(\cdot)$ and obtained $LL^h(\mathcal{N}(0)^h, \hat{\phi}^h, \hat{\rho}^h, \hat{\delta}^l, \hat{\boldsymbol{\lambda}}^h)$. The likelihood ratio test statistics are given as follows:

$$LR = -2(LL^h(\mathcal{N}(0)^h, \hat{\phi}^h, \hat{\rho}^h, \hat{\delta}^l, \hat{\lambda}^h) - LL^h(\hat{\theta}^h)).$$

This statistic follows a chi-square distribution with $m = 1$ degrees of freedom.

The results of the two likelihood ratio tests are presented in Tables 6 (Model IV) and 7 (Model V). The results of the general test are given in the upper panel of Tables 6 and 7, whereas the results of the specific test are given in the lower panel of Tables 6 and 7. In fact, the results presented in both tables are more extensive than what we have illustrated above. For the high WMC group, in addition to the upper half, we also estimate the $\boldsymbol{\theta}$ of the top one-third (WMC $> P_{67}$) and the top one-fourth (WMC $> P_{75}$); similarly, for the low group, in addition to the lower half, the $\boldsymbol{\theta}$ of the bottom one-third (WMC $< P_{33}$) and one-fourth (WMC $< P_{25}$) are also estimated. Therefore, what is presented in Tables 6 and 7 are the likelihood ratio tests of all heterogeneous pairs, i.e., pairs involving one high WMC group and one low WMC group. Here we do not consider comparisons based on the homogeneous pairs since their observations are partially overlapping. Altogether we have 9 ($3 \times 3$) heterogeneous pairs (the cross product of three high WMC groups and three low WMC groups). The results can be easily arrayed into a three-by-three matrix, as shown in both the upper and lower panel of Tables 6 and 7. In addition, our LR tests were corrected for multiple comparisons, in our case, nine comparisons, by controlling the *family-wise error rate* (Šidák, 1967).

Let us first look at the general test (the upper panel of Tables 6 and 7). An interesting pattern immediately standing out is that the likelihood ratio test is statistically significant over all pairs. This can be clearly seen from the asterisks, denoting the significance level, which are presented beside the $\chi^2$ statistics. The only one exception is the pair 'WMC $> P_{75}$' vs. 'WMC < mean' under Model V (Table 7). Therefore, by and large, these results support our working memory hypothesis for individual differences in learning, i.e., the learning behavior of subjects is different between low and high WMC groups. This seems to confirm the earlier conjecture made by Chen et al. (2014) that WMC not only affects subjects' performance, but may also have effects on their learning schemes.

As to the specific test of the counterfactual thinking ability ($\delta$), we found that the $\hat{\delta}$ of the top one-half of subjects is significantly higher than those of the bottom one-third and the bottom one-fourth of subjects, but there is no evidence showing that it is different from the $\hat{\delta}$ of the bottom one-half of subjects (this result is consistent for both Models IV and V). The increase in WMC does not seem to be enhancing; in fact, the $\hat{\delta}$ of the top one-third and the top one-fourth of subjects exhibited no significant difference to any group of low-WMC subjects (the only one exception is the pair 'WMC > P67' vs. 'WNC < P25' in Model V).

Hence, our second hypothesis, the working memory hypothesis for individual differences in counterfactual thinking ability, is not well supported by our data, and a monotonically increasing relation between WMC and counterfactual thinking ability can be rejected outright. The results of the two tables together seem to suggest that working memory capacity can have an effect on counterfactual thinking ability ($\delta$) only when working memory capacity is low, down to the bottom one-fourth or the bottom one-third. When WMC gets closer to its average or is above its

Table 6
LR test for the significance of differences in parameter estimates: Model IV.

| Parameters | WMC < mean | WMC < $P_{33}$ | WMC < $P_{25}$ |
|---|---|---|---|
| | *General comparisons:* $\hat{\theta}^l \rightarrow LL^h(\cdot)$ | | |
| WMC > $P_{75}$ | 64.594$^{***}$(<0.001) | 79.970$^{***}$(<0.001) | 96.288$^{***}$(<0.001) |
| WMC > $P_{67}$ | 54.168$^{***}$(<0.001) | 72.182$^{***}$(<0.001) | 87.928$^{***}$ (<0.001) |
| WMC > mean | 53.422$^{***}$(<0.001) | 73.158$^{***}$(<0.001) | 97.536$^{***}$(<0.001) |
| | *Single parameter comparisons:* $\hat{\delta}^l \rightarrow LL^h(\cdot)$ | | |
| | ($\hat{\delta} = 0.472$) | ($\hat{\delta} = 0.405$) | ($\hat{\delta} = 0.314$) |
| WMC > $P_{75}$($\hat{\delta} = 0.411$) | 1.178(0.278) | 0.188(0.665) | 0.928(0.335) |
| WMC > $P_{67}$($\hat{\delta} = 0.484$) | 0.016(0.899) | 0.772(0.380) | 3.978(0.046) |
| WMC > mean($\hat{\delta} = 0.578$) | 2.472(0.116) | 7.118$^{*}$(0.008) | 17.262$^{***}$(<0.001) |

The test statistic $\chi^2$ is shown in the table. The $p$ value of each $\chi^2$ is also shown inside the parentheses below. We control the family-wise error rate (FWER), denoted as $\alpha$, using the method proposed by Zbyněk Šidák (Šidák, 1967). Given $k$ independent comparions ($k = 9$), each null hypothesis is rejected when the $p$-value is lower than $1 - (1 - \alpha)^{1/k}$. The significance level $\alpha$ are set at 5%, and they are denoted as $^{**}$. The critical values for general comparisons are $\chi^2_{0.9884}(11) = 24.271$, $\chi^2_{0.9943}(11) = 26.386$, and $\chi^2_{0.9989}(11) = 30.964$. The critical values for single parameter comparisons are $\chi^2_{0.9884}(1) = 6.365$, $\chi^2_{0.9943}(1) = 7.648$, and $\chi^2_{0.9989}(1) = 10.624$.

$^{*}$ The significance level $\alpha$ are set at 10%.

$^{***}$ The significance level $\alpha$ are set at 1%.

Table 7
LR test for the significance of differences in parameter estimates: Model V.

| Parameters | WMC < mean | WMC < $P_{33}$ | WMC < $P_{25}$ |
|---|---|---|---|
| | *General comparisons:* $\hat{\theta}^l \rightarrow LL^h(\cdot)$ | | |
| WMC > $P_{75}$ | 10.194(0.070) | 26.540***(<0.001) | 48.906***(<0.001) |
| WMC > $P_{67}$ | 17.028*(0.004) | 37.292***(<0.001) | 63.370***(<0.001) |
| WMC > mean | 16.828**(<0.001) | 38.232***(<0.001) | 66.324***(<0.001) |
| | *Single parameter comparisons:* $\hat{\delta}^l \rightarrow LL^h(\cdot)$ | | |
| | ($\hat{\delta} = 0.4894$) | ($\hat{\delta} = 0.3998$) | ($\hat{\delta} = 0.2774$) |
| WMC > $P_{75}(\hat{\delta} = 0.4635)$ | 0.114(0.736) | 0.688(0.407) | 5.866(0.015) |
| WMC > $P_{67}(\hat{\delta} = 0.5582)$ | 0.982(0.322) | 5.212(0.022) | 16.370***(<0.001) |
| WMC > mean($\hat{\delta} = 0.5877$) | 2.702(0.100) | 9.846**(0.002) | 26.724***(<0.001) |

The test statistic $\chi^2$ is shown in the table. The $p$ value of each $\chi^2$ is also shown inside the parentheses below. We control the family-wise error rate (FWER), denoted as $\alpha$, using the method proposed by Zbyněk Šidák (Šidák, 1967). Given $k$ independent comparions ($k = 9$), each null hypothesis is rejected when the $p$-value is lower than $1 - (1 - \alpha)^{1/k}$. The critical values for general comparisons are $\chi^2_{0.9884}(5) = 14.718$, $\chi^2_{0.9943}(5) = 16.445$, and $\chi^2_{0.9989}(5) = 20.261$. The critical values for single parameter comparisons are $\chi^2_{0.9884}(1) = 6.365$, $\chi^2_{0.9943}(1) = 7.648$, and $\chi^2_{0.9989}(1) = 10.624$.
* The significance level $\alpha$ are set at 10%.
** The significance level $\alpha$ are set at 5%..
*** The significance level $\alpha$ are set at 1%.

average, the difference becomes less certain; specifically, the subjects with superb WMC (the top one-fourth) do not demonstrate their superiority in terms of $\delta$. This latter evidence may suggest that the counterfactual thinking ability may become flat after WMC increases to its average or even a little before its average. This tendency is what is numerically presented in both tables. For example, in Tables 6 and 7, $\hat{\delta}$ begins with 0.31 (0.28) when only the bottom one-fourth of subjects are considered and increases all the way up to 0.57 (0.59) when only the top one-half of subjects are considered. It then declines slightly to 0.48 (0.56) and falls further to 0.41 (0.46) when the subjects considered are restricted to the top one-third and the top one-fourth.

## 5. Discussion

This paper, to the best of our knowledge, is the first study to provide a detailed account of the inferred individual differences in EWA learning models in light of subjects' working memory capacity. In this section, we want to highlight the three key results presented in the previous section; there are three aspects, namely,

- the validity of the constrained version of generalized reinforcement learning (Section 5.1),
- the working memory hypothesis for individual differences in learning (Section 5.2), and
- the working memory hypothesis for individual differences in counterfactual thinking ability (Section 5.3).

### 5.1. Cognitively constrained models of reinforcement learning

First, we have examined the generalized reinforcement learning models without the constraint of cognitive capacity (Models I-III) and with the constraint of cognitive capacity (Models IV-V). From the behavior of the predicted choice probabilities and the range of some estimates, such as $\hat{\phi}$, it is found that the constrained version of generalized reinforcement learning performs more reasonably than the unconstrained version. As we mention in Section 1.2, the problem of the unconstrained version was acknowledged in the economics literature (Brenner & Vriend, 2006; Camerer & Ho, 1999) but only implicitly. However, the psychological underpinning of this problem has been well developed with the coined term 'magic number' (Mathy & Feldman, 2012). Recently, Collins and Koechlin (2012) have found that the model that best fits human data is endowed with a monitoring capacity of three or four task sets, suggesting that working memory is limited to three or four concurrent behavioral strategies. Hence, our finding in this regard is relevant and contributes to this line of research.

Despite this being the case, a point which has been made earlier (Section 3.4.2) and which needs to be reemphasized here is that what concerns us here is not just size per se, but to a somewhat greater extent the associated structure. What seems to be more impressive about the level-$k$ reasoning is that the six intervals are updated and may shrink over time, which provides additional flexibility for Models IV-V, and may also be a better description of what subjects are actually doing. As we move from the number reinforcement models to the level reinforcement models, subjects are assumed to redefine learning objects that are being evaluated and updated as consisting of a few sophisticated rules. Under such circumstances, the reinforcement mechanism works at a more abstract level (the choice of reasoning depths) instead of at a primitive stage (the choice of guessing numbers).

Alternatively put, reducing dimensionality comes along with some forms of abstraction from the original

unprocessed representation of the problem. In the literature on reinforcement learning when it serves as an algorithm to solve a computational problem, recent studies in hierarchical reinforcement learning (HRL) share a similar idea (see Botvinick, 2012 a review). The HRL framework was proposed to resolve the curse of dimensionality which causes the deterioration of efficiency. It allows the agent to select temporally abstract actions, and therefore reduces the number of alternatives the agent has to learn about. How the relevant forms of abstraction are initially acquired or learned is the central issue in this line of research, while the application of level-$k$ models could be regarded as a natural abstraction in our guessing game.

### 5.2. WMH for individual differences in learning

Second, given the above result, our subsequent efforts were directed toward the constrained version of the EWA learning model (Models IV and V). As introduced in Section 2.1, EWA learning models have three kinds of cognitive constructs, namely, the memory-related parameters ($\phi, N(0)$ and $\rho$), choice intensity ($\lambda$), and counterfactual thinking ability ($\delta$). At a finer level, one may want to know whether each of these parameters can be related to WMC in a certain way. However, as argued in Section 2.2, a good maintained hypothesis can only be found for the relationship between WMC and $\delta$. Nonetheless, we are still interested in knowing whether WMC has an effect on (generalized) reinforcement learning as a whole. For that purpose, we formed the working memory hypothesis for individual differences in learning and found that the hypothesis can be well supported by our data (see Tables 6 and 7).

This result is related to the recent efforts to incorporate the working memory component into reinforcement learning (Collins & Frank, 2012; Dolan & Dayan, 2013). Collins and Frank (2012) include the capacity-limited working memory component in a simple reinforcement learning system, and show that fitting the data with a reinforcement learning model alone can cause the estimated learning rate parameters to be misleading, because it will capture the effect introduced by working memory capacity. In this regard, our paper provides additional evidence showing that this influence of working memory capacity may be applicable to generalized versions of reinforcement learning.

Furthermore, as motivated by Collins and Frank (2012) and their model-based learning, we can go further to hypothesize that subjects with different working memory capacity can have different geometries of strategy space, for example, hierarchical ones. It is unlikely that all humans homogeneously resolve the size problem by using the same granulation. Subjects with different working memory capacity may also apply different forms of granulation. So far, we have known very little about these individual differences, particularly in the context of economic decision making. Currently, the kinds of reinforcement learning models used in economics have a flat structure with no restrictions on the span. When the number of options is far beyond the 'magic number', what the appropriate representation of the strategy space that allows reinforcement learning models to be effectively applied actually constitutes an issue for further study.

### 5.3. WMH for individual differences in CT ability

Third, by estimating the constrained EWA learning models, we have also pinpointed the effect of working memory capacity on various cognitive constructs. We have found that subjects with high WMC tend to learn faster (a lower $N(0)$), are more sensitive to recent experience (a lower $\rho$), and have a slower memory decay rate (a larger $\phi$). What interests us most is the second maintained hypothesis, namely, the working memory hypothesis for individual differences in counterfactual thinking ability. Although it is only weakly supported by the data, the positive effect of working memory capacity on counterfactual thinking ability has been found in a number of studies (Byrne, 2016; Camille et al., 2004; Goldinger et al., 2003; Kulakova & Nieuwland, 2016).

We also understand that research on the biological and neural underpinnings of reinforcement learning has already been undergoing a thorough understanding of the dopamine neural system. By hypothesizing that dopamine neurons encode reward prediction errors, a hypothesis widely known as the *reward prediction error hypothesis* has been developed for decades (Bayer & Glimcher, 2005; Glimcher, 2010; Schultz & Romo, 1990; Schultz, Dayan, & Montague, 1997; Zhu, 2011). Recently, by directly measuring dopamine release in the human striatum, Kishida et al. (2016) have found that dopamine levels reflect a combination of reward prediction errors and counterfactual prediction errors. To account for this result, their proposed computation model indicates that dopamine neurons compute not only experienced rewards and losses but also the rewards and losses that might have been experienced if the alternative had been taken. Therefore, their finding extends the long-held reward prediction error hypothesis and directly implies that dopamine also encodes counterfactual prediction errors, a key variable posited by models of generalized reinforcement learning addressed in this paper (see also Montague, King-Casas, & Cohen, 2006).

Within this research circle, our result on weakly supporting the positive relation between counterfactual thinking ability ($\delta$) and working memory capacity may stimulate new research questions. In particular, it places itself at an initial stage of a more integrated framework which overarches psychological studies of counterfactual thinking, neuroscientific studies of generalized reinforcement learning mechanisms, and behavioral studies of strategic decision-making in economic games. Psychologists have conducted various tasks to measure counterfactual thinking ability; however, at this point, it is not clear how $\delta$ as part of the generalized reinforcement learning can be related to those

tasks. A follow-up study would be to carry out an independent counterfactual thinking test for the subjects, and *directly* examine the relationship between WMC and counterfactual thinking ability. Then we could use this result to shed light on the connection between WMC and $\delta$ derived from experimental games. This line of research will help us to gain a better understanding of the cognitive role of $\delta$.

## 6. Conclusions

In conclusion, the main findings and contribution of the paper can be summarized as follows. First, we have shown that the generalized reinforcement learning model may work reasonably well when the number of choices (armed bandits) is constrained by our working memory capacity, such as Miller's seven. As we have shown from the dynamics (evolution) of choice probabilities, the level reinforcement model (with six alternatives only) behaves very well compared to the number reinforcement model (with an overwhelming 101 alternatives). This finding may not surprise psychologists, but there has been a general lack of awareness among economists. Second, while restricted to the level reinforcement model, the working memory hypothesis for individual differences in learning is well supported by our data from the beauty contest games (guessing games). Even after the multiple comparison correction, 17 out of 18 pairs of heterogeneous groups in WMC exhibit significant differences in learning. This result together with that of Chen et al. (2014) shows that the observed behavioral difference among subjects with different WMC is sustained because they actually learned in a different way. Third, as to the working memory hypothesis for individual differences in counterfactual thinking ability, we find that this hypothesis is also supported in the following sense: the parameter corresponding to counterfactual thinking ability increases with working memory capacity at its initial level, but then flattens out at its middle and high levels. Each of the results has been discussed, and possible directions for further studies provided.

## Acknowledgements

Earlier versions of this paper were presented at the 2013 Regional Economic Science Association (ESA) Conference, Santa Cruz, California, October 24–26, 2013, the NeuroPsychoEconomics Conference, Munich, Germany, May 29–30, 2014, the 21st International Conference on Computing in Economics and Finance, Taipei, June 20–22, 2015, and the North America Conference of the Chinese Economists Society (CES), Sacramento, California, USA, April 2–3, 2016. The authors benefited significantly from the discussions with conference participants. This version has been substantially revised in light of two anonymous referees' very painstaking reviews, for which we are most grateful. Research support in the form of the Ministry of Science and Technology – Taiwan (MOST) Grant, MOST 103-2410-H-004-009-MY3, is gratefully acknowledged.

## Appendix A. Working memory test

*Backward Digital Span Test (Dspan)*. This task was to recall a set of digits in reverse order. Following a fixation cross presented for 1 s, a set of 4–8 digits were displayed one by one, for 1 s each. After that, subjects were required to enter this set of digits in reverse order without time constraints. There were 15 trials in total, with 3 at each set size.

*Spatial Short-term Memory Test (SSTM)*. The subjects were required to memorize the location of a set of dots in a $10 \times 10$ grid. This task started with a fixation cross for 1 s and the grid was shown. There were 2–6 solid dots that appeared, one by one, in individual cells, for 900 ms each. The interstimulus interval was 100 ms. The subjects were instructed to remember the spatial relation of the dots instead of the absolute position of each dot. After presenting all of the dots, the subjects were asked to replicate the pattern of dots. There were 30 trials, with 6 at each set size.

*Memory Updating Test (MU)*. This task was to encode a set of digits, each presented sequentially in a set of frames, and then to update these digits by arithmetical operations. In each trial, the subjects were presented with 3–5 frames containing to-be-remembered digits in each. Each trial was initialized by a keypress and then the initial digits were displayed one by one, for 1 s each. After that, 2–6 arithmetical operations, such as "+3" or "−1", were displayed in individual frames one by one for 1.3 s each and followed by a 250-ms blank interval. Subjects were required to apply these operations to the digits that they currently remembered in that particular frame and to update the content with the result. There were 15 trials in total.

*Sentence Span Test (SentSpan)*. On each trial, an alternating sequence of Chinese sentences and to-be-remembered consonants was presented. The subjects had to judge the meaningfulness of the sentences and to remember the following consonants for later serial recall. The sentences were composed of 17 Chinese characters. For example, a meaningful sentence might be *I went out without taking any money, but fortunately I ran into an old friend who helped me out*. By replacing *fortunately* with *unfortunately* we obtained the meaningless counterpart of this sentence. Following a fixation cross presented for 1.5 s, subjects saw the first sentence appear on the screen. It disappeared either when subjects gave a response or after the maximal response time of 5 s had elapsed. The subjects were instructed to use the "/" and "z" keys to make *Yes, this is correct* and *No, this is not correct* responses, respectively. After a judgment was made on a sentence, a consonant was presented for 1 s. After the consonant disappeared, the next sentence appeared. The list length, defined as the number of sentences and letters needed to be judged and remembered, ranged from 4 to 8. There were 15 trials in total, with 3 trials per list length.

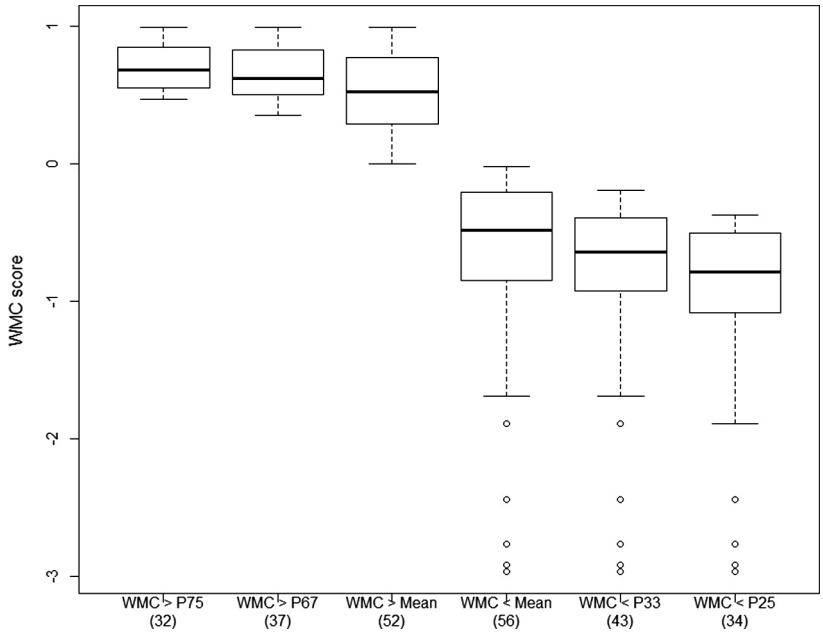



Fig. B.4. WMC score distribution of each group. The above figure shows the box-whisker plot of various subgroups of subjects in terms of WMC. From the leftmost to the rightmost are the group "the top one-fourth", "the top one-third", "above average", "below average", "the bottom one-third", and "the bottom one-fourth". What is shown inside the parentheses is the number of subjects in the respective group.

*Operation Span Test (OS)*. This task was almost the same as the SentSpan task except that the subjects had to judge the correctness of the arithmetic equations (e.g., $3 + 2 = 5$). A minor difference was that the maximum response time for the equation was set to 3 s due to the simplicity of this processing task.

## Appendix B. Distribution of various subgroups

In Fig. B.4, we provide the distribution (the box-whisker plot) of the various subgroups of subjects considered in this study. From what has been shown in the figure, we can see that the high and low WMC groups are clustered in discernible disparate levels.

## References


Arthur, B. (1993). On designing economic agents that behave like human agents. *Journal of Evolutionary Economics, 3*(1), 1–22.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron, 47*(1), 129–141.

Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Current Opinion in Neurobiology, 22*, 956–962.

Brañas-Garza, P., García-Muñoz, T., & González, R. H. (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization, 83*(2), 254–260.

Brenner, T., & Vriend, N. J. (2006). On the behavior of proposers in ultimatum games. *Journal of Economic Behavior & Organization, 61*(4), 617–631.

Brock, W. A., & Hommes, C. H. (1997). A rational route to randomness. *Econometrica: Journal of the Econometric Society*, 1059–1095.

Broseta, B. (2000). Adaptive learning and equilibrium selection in experimental coordination games: An ARCH(1) approach. *Games and Economic Behavior, 32*(1), 25–50.

Burnham, T. C., Cesarini, D., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Higher cognitive ability is associated with lower entries in a p-beauty contest. *Journal of Economic Behavior & Organization, 72*(1), 171–175, October.

Bush, R. R., & Mosteller, F. (1955). Stochastic models for learning. New York: John Wiley & Sons.

Byrne, R. M. (2005). The rational imagination: How people create alternatives to reality. Cambridge, MA: MIT Press.

Byrne, R. M. (2016). Counterfactual thought. *Annual Review of Psychology, 67*, 135–157.

Camerer, C. F., & Ho, T.-H. (1998). Experience-weighted attraction learning in coordination games: Probability rules, heterogeneity, and time-variation. *Journal of Mathematical Psychology, 42*(2), 305–326.

Camerer, C. F., & Ho, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica, 67*(4), 827–874.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2002). Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory, 104*(1), 137–188.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., & Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science, 304*, 1167–1170.

Casari, M., Ham, J. C., & Kagel, J. H. (2007). Selection bias, demographic effects, and ability effects in common value auction experiments. *The American Economic Review*, 1278–1304.

Chen, S.-H., Du, Y.-R., & Yang, L.-X. (2014). Cognitive capacity and cognitive hierarchy: A study based on beauty contest experiments. *Journal of Economic Interaction and Coordination, 9*, 69–105.

Chen, S.-H., & Hsieh, Y.-L. (2011). Reinforcement learning in experimental asset markets. *Eastern Economic Journal, 37*(1), 109–133.

Chen, Y., & Khoroshilov, Y. (2003). Learning under limited information. *Games and Economic Behavior, 44*, 1–25.

Cheung, Y.-W., & Friedman, D. (1997). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior, 19*(1), 46–76.

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience, 35*, 1024–1035.

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLoS Biology, 10* (3), e1001293.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 871–885.

Crawford, V. (1995). Adaptive dynamics in coordination games. *Econometrica, 63*, 103–143.

Cross, J. (1973). A stochastic learning model of economic behavior. *Quarterly Journal of Economics, 87*(2), 239–266.

Devetag, G., & Warglien, M. (2003). Games and phone numbers: Do short-term memory bounds affect strategic behavior? *Journal of Economic Psychology, 24*(2), 189–202.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron, 80*(2), 312–325.

Duffy, J., & Nagel, R. (1997). On the robustness of behaviour in experimental "beauty contest games. *Economic Journal, 107*(445), 1684–1700.

Ferguson, H. J. (2012). Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. *The Quarterly Journal of Experimental Psychology, 65*, 939–961.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Gill, D., & Prowse, V. (2012). Cognitive ability and learning to play equilibrium: A level-k analysis, working paper.

Glimcher, P. W. (2010). Foundations of neuroeconomic analysis. Oxford University Press.

Goldinger, S. D., Kleider, H. M., Azuma, T., & Beike, D. R. (2003). "blaming the victim" under memory load. *Psychological Science, 14*(1), 81–85.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*(2), 243–266.

Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review, 88*(4), 947–969.

Hommes, C. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control, 35* (1), 1–24.

Ho, T. H., Wang, X., & Camerer, C. F. (2008). Individual differences in EWA learning with partial payoff information. *The Economic Journal, 118*(525), 37–59.

Iyengar, S., Huberman, G., & Jiang, W. (2004). How much choice is too much? Contributions to 401(k) retirement plans. In O. Mitchell & S. Utkus (Eds.), *Pension design and structure: Lessons from behavioral finance* (pp. 83–95). Oxford University Press.

Iyengar, S., & Lepper, M. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79*(6), 995–1006.

Kishida, K. T., Saez, I., Lohrenz, T., Witcher, M. R., Laxton, A. W., Tatter, S. B., ... Montague, P. R. (2016). Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. *Proceedings of the National Academy of Sciences, 113*(1), 200–205.

Kulakova, E., & Nieuwland, M. S. (2016). Understanding counterfactuality: A review of experimental evidence for the dual meaning of counterfactuals. *Language and Linguistics Compass, 10*(2), 49–65.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. (2010). A working memory test battery for matlab. *Behavior Research Methods, 42*(2), 571–585.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition, 122*, 346–362.

Miller, G. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Montague, P. R., King-Casas, B., & Cohen, J. D. (2006). Imaging valuation models in human choice. *Annual Review of Neuroscience, 29*, 417–448.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review, 85*(5), 1313–1326.

Roth, A., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior, 8*, 164–212.

Rydval, O., Ortmann, A., & Ostatnicky, M. (2009). Three very simple games and what it takes to solve them. *Journal of Economic Behavior & Organization, 72*(1), 589–601, October.

Sarin, R., & Vahid, F. (2004). Strategy similarity and coordination. *The Economic Journal, 114*, 506–527.

Scheibehenne, B., Greifeneder, R., & Todd, P. (2010). Can there ever be too many options? A metaanalytic review of choice overload. *Journal of Consumer Research, 37*(3), 409–425.

Schnusenberg, O., & Gallo, A. (2011). On cognitive ability and learning in a beauty contest. *Journal for Economic Educators, 11*(1), 13–24.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599.

Schultz, W., & Romo, R. (1990). Dopamine neurons of the monkey midbrain: Contingencies of responses to stimuli eliciting immediate behavioral reactions. *Journal of Neurophysiology, 63*(3), 607–624.

Šidák, Z. K. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62*(318), 626–633.

Stahl, D. O. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior, 16*(2), 303–330.

Stahl, D. O. (1998). Is step-*j* thinking an arbitrary modelling restriction or a fact of human nature? *Journal of Economic Behavior & Organization, 37*(1), 33–51.

Stahl, D. (2000). Rule learning in symmetric normal-form games: Theory and evidence. *Games and Economic Behavior, 32*(1), 105–138.

Thaler, R., & Sunstein, C. (2008). Nudge: Improving decisions about health, wealth, and happiness. Penguin Books.

Thorndike, E. (1911). *Animal intelligence: Experimental studies*. *The animal behaviour series*. Macmillan.

Urrutia, M., de Vega, M., & Bastiaansen, M. (2012). Understanding counterfactuals in discourse modulates ERP and oscillatory gamma rhythms in the EEG. *Brain Research, 1455*, 40–55.

Zhu, L. (2011). *Understanding neural mechanisms of strategic learning: Correlates, causality, and applications*. University of Illinois at Urbana-Champaign.